```
US-08-653-294-12 x HSHLABD1

Align seg 1/1 to: HSHLABD1    from: 1 to: 250

      1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
        |||||:::|||||||||||||||||||
    202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABF1

seq_documentation_block:
LOCUS       HSHLABF1      250 bp    DNA           PRI     25-MAR-1997
DEFINITION  Human cell line THAI DCH028 MHC class I HLA-B gene (allele
            HLA-B*51V), exon 2.
ACCESSION   U90613
VERSION     U90613.1  GI:1906033
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       B*51V alleles
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       Direct Submission
JOURNAL     Submitted (25-FEB-1997) Transfusion Medicine, Faculty of Medicine,
            Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
            Thailand
FEATURES            Location/Qualifiers
     source         1..250
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /map="6p21"
                    /haplotype="(a)HLA-A1,B51V,Cw14,DR6(DR52),DQ1/(b)A2,B77,
                    Cw8N,DR4(DR53),DQ4"
                    /cell_type="lymphoblastoid"
                    /cell_line="THAI DCH028"
     exon           1..250
                    /gene="HLA-B"
                    /note="Allele: HLA-B*51V; similar to exon 2 of B*5105"
                    /number=2
BASE COUNT     56 a     82 c     80 g     32 t
ORIGIN

alignment_scores:
       Quality:   44.00               Length:      10
         Ratio:    4.400               Gaps:         0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABF1

Align seg 1/1 to: HSHLABF1    from: 1 to: 250

      1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
        |||||:::|||||||||||||||||||
    202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABG1

seq_documentation_block:
LOCUS       HSHLABG1      250 bp    DNA           PRI     25-MAR-1997
DEFINITION  Human cell line THAI DCH011 MHC class I HLA-B gene (allele
            HLA-B*51V), exon 2.
ACCESSION   U90615
```

```
     exon           1..250
                    /gene="HLA-B"
                    /number=2
                    /product="MHC class I HLA-B"
BASE COUNT     55 a     83 c     80 g     32 t
ORIGIN

alignment_scores:
       Quality:   44.00               Length:      10
         Ratio:    4.400               Gaps:         0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABB1

Align seg 1/1 to: HSHLABB1    from: 1 to: 250

      1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
        |||||:::|||||||||||||||||||
    202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABD1

seq_documentation_block:
LOCUS       HSHLABD1      250 bp    DNA           PRI     22-MAR-1997
DEFINITION  Human cell line THAI DCH010 MHC class I HLA-B gene (allele
            HLA-B*51V), exon 2.
ACCESSION   U90611
VERSION     U90611.1  GI:1905865
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       B*51V alleles
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       Direct Submission
JOURNAL     Submitted (25-FEB-1997) Transfusion Medicine, Faculty of Medicine,
            Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
            Thailand
FEATURES            Location/Qualifiers
     source         1..250
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /map="6p21"
                    /haplotype="(b)HLA-A2,B77,Cw8N,DR12(DR52),DQ7/(c)A11.1,
                    B51V,Cw14,DR12(DR52),DQ1"
                    /cell_type="lymphoblastoid"
                    /cell_line="THAI DCH010"
     exon           1..250
                    /gene="HLA-B"
                    /note="Allele: HLA-B*51V; similar to exon 2 of B*5105"
                    /number=2
                    /product="MHC class I HLA-B"
BASE COUNT     56 a     82 c     80 g     32 t
ORIGIN

alignment_scores:
       Quality:   44.00               Length:      10
         Ratio:    4.400               Gaps:         0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
```

```
VERSION     U90615.1  GI:1906037
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       B*51V alleles
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Siriboonrit,U., Rungroung,E. and Bejchandra,S.
TITLE       Direct Submission
JOURNAL     Submitted (25-FEB-1997) Transfusion Medicine, Faculty of Medicine,
            Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
            Thailand
FEATURES             Location/Qualifiers
     source          1..250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /haplotype="(a)HLA-A11.1,B51V,Cw14,DR12(DR52),DQ7/(b)A24,
                     B77,Cw8N,DR12(DR52)DQ7"
                     /cell_type="lymphoblastoid"
                     /cell_line="THAI DCH011"
     exon            1..250
                     /gene="HLA-B"
                     /number=2
                     /note="Allele: HLA-B*51V; similar to exon 2 of B*5105"
BASE COUNT       56 a       82 c       80 g       32 t
ORIGIN

alignment_scores:
        Quality:     44.00                  Length:      10
        Ratio:        4.400                  Gaps:         0
Percent Similarity: 100.000      Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABG1          ..

Align seg 1/1  to: HSHLABG1  from: 1  to: 250

          1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
            ||||||:::::||||||||||||||||||||
        202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABI1

seq_documentation_block:
LOCUS       HSHLABI1      250 bp    DNA        PRI       22-MAR-1997
DEFINITION  Human cell line THAI DCH011 MHC class I HLA-B gene (allele
            HLA-B*1513), exon 2.
ACCESSION   U90422
VERSION     U90422.1  GI:1905834
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Rungroung,E., Bejchandra,S., Blasczyk,R. and Grosse-Wilde,H.
TITLE       B15 alleles (B*1513)
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Rungroung,E., Bejchandra,S., Blasczyk,R. and Grosse-Wilde,H.
```

```
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-1997) Transfusion Medicine, Faculty of Medicine,
            Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
            Thailand
FEATURES             Location/Qualifiers
     source          1..250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="VI"
                     /map="6p21"
                     /cell_type="lymphoblastoid"
                     /cell_line="THAI DCH011"
     exon            1..250
                     /gene="HLA-B"
                     /note="Allele: HLA-B*1513"
                     /number=2
                     /product="MHC class I HLA-B"
BASE COUNT       55 a       83 c       80 g       32 t
ORIGIN

alignment_scores:
        Quality:     44.00                  Length:      10
        Ratio:        4.400                  Gaps:         0
Percent Similarity: 100.000      Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABI1          ..

Align seg 1/1  to: HSHLABI1  from: 1  to: 250

          1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
            |||||||::::|||||||||||||||||||
        202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABJ1

seq_documentation_block:
LOCUS       HSHLABJ1      250 bp    DNA        PRI       22-MAR-1997
DEFINITION  Human cell line THAI DCH028 MHC class I HLA-B gene (allele
            HLA-B*1513), exon 2.
ACCESSION   U90424
VERSION     U90424.1  GI:1905838
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Rungroung,E. and Bejchandra,S.
REFERENCE   2 (bases 1 to 250)
AUTHORS     Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
            Rungroung,E. and Bejchandra,S.
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-1997) Transfusion Medicine, Faculty of Medicine,
            Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
            Thailand
FEATURES             Location/Qualifiers
     source          1..250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="VI"
                     /map="6p21"
                     /cell_type="lymphoblastoid"
                     /cell_line="THAI DCH028"
     exon            1..250
                     /gene="HLA-B"
                     /note="Allele: HLA-B*1513"
                     /number=2
```

```
                                                        |||||:::||||||||||||||||||||||
                                                    202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr4:AF022159

seq_documentation_block:
LOCUS        AF022159       255 bp    DNA              PRI       05-JAN-1999
DEFINITION   Homo sapiens isolate 026 MHC class I antigen HLA-H (HLA-H)
             pseudogene, partial sequence.
ACCESSION    AF022159
VERSION      AF022159.1  GI:2655062
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 255)
  AUTHORS    Grimsley,C., Mather,K.A. and Ober,C.
  TITLE      HLA-H: a pseudogene with increased variation due to balancing
             selection at neighboring loci
  JOURNAL    Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
  MEDLINE    99083426
REFERENCE    2  (bases 1 to 255)
  AUTHORS    Grimsley,C., Mather,K.A. and Ober,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-SEP-1997) Fred Hutchinson Cancer Research Center,
             1100 Fairview Ave. N., M374, Seattle, WA 98109, USA
FEATURES             Location/Qualifiers
     source          1..255
                     /organism="Homo sapiens"
                     /isolate="026"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /note="African-American individual"
     exon            1..255
                     /gene="HLA-H"
                     /pseudo
                     /number=2
     CDS             <1..>255
                     /gene="HLA-H"
                     /note="MHC class I antigen HLA-H"
                     /codon_start=1
                     /pseudo
     gene            <1..>255
                     /gene="HLA-H"
                     /pseudo
BASE COUNT      50 a     81 c     86 g     38 t
ORIGIN

alignment_scores:
     Quality:   44.00         Length:     10
     Ratio:      4.400        Gaps:        0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x AF022159

Align seg 1/1  to: AF022159  from: 1  to: 255

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      |||||:::|||||||||||||||||||||||
  223 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 252

seq_name: gb_pr4:AF022160

seq_documentation_block:
LOCUS        AF022160       259 bp    DNA              PRI       05-JAN-1999
DEFINITION   Homo sapiens isolate 034 MHC class I antigen HLA-H (HLA-H)
             pseudogene, partial sequence.
ACCESSION    AF022160
VERSION      AF022160.1  GI:2655063
```

```
BASE COUNT      55 a     83 c     80 g     32 t
ORIGIN

alignment_scores:
     Quality:   44.00         Length:     10
     Ratio:      4.400        Gaps:        0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABJ1

Align seg 1/1  to: HSHLABJ1  from: 1  to: 250

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      |||||:::|||||||||||||||||||||||
  202 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 231

seq_name: gb_pr2:HSHLABT1

seq_documentation_block:
LOCUS        HSHLABT1       250 bp    DNA              PRI       22-MAR-1997
DEFINITION   Human cell line THAI DCH009 MHC class I HLA-B gene (allele
             HLA-B*1513), exon 2.
ACCESSION    U90418
VERSION      U90418.1  GI:1905826
SEGMENT      1 of 2
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 250)
  AUTHORS    Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
             Rungroung,E., Bejchandra,S., Juji,T. and Tokunaga,K.
  TITLE      B15 alleles (B*1513)
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 250)
  AUTHORS    Chandanayingyong,D., Sirikong,M., Longta,K., Srinak,D.,
             Rungroung,E., Bejchandra,S., Juji,T. and Tokunaga,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (23-FEB-1997) Transfusion Medicine, Faculty of Medicine,
             Siriraj Hospital, Mahidol University, Prannok Road, Bangkok 10700,
             Thailand
FEATURES             Location/Qualifiers
     source          1..250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="VI"
                     /map="6p21"
                     /cell_type="lymphoblastoid"
                     /cell_line="THAI DCH009"
     exon            1..250
                     /gene="HLA-B"
                     /note="Allele: HLA-B*1513"
                     /number=2
BASE COUNT      55 a     83 c     80 g     32 t
ORIGIN                   /product="MHC class I HLA-B"

alignment_scores:
     Quality:   44.00         Length:     10
     Ratio:      4.400        Gaps:        0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x HSHLABT1

Align seg 1/1  to: HSHLABT1  from: 1  to: 250

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
```

```
KEYWORDS     human.
SOURCE       Homo sapiens
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 259)
AUTHORS      Grimsley,C., Mather,K.A. and Ober,C.
TITLE        HLA-H: a pseudogene with increased variation due to balancing
             selection at neighboring loci
JOURNAL      Mol. Biol. Evol. 15 (12), 1581-1588 (1998)
MEDLINE      99083426
REFERENCE    2 (bases 1 to 259)
AUTHORS      Grimsley,C., Mather,K.A. and Ober,C.
TITLE        Direct Submission
JOURNAL      Submitted (03-SEP-1997) Fred Hutchinson Cancer Research Center,
             1100 Fairview Ave. N., M374, Seattle, WA 98109, USA
FEATURES             Location/Qualifiers
     source          1..259
                     /organism="Homo sapiens"
                     /isolate="034"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /note="African-American individual"
     exon            1..259
                     /gene="HLA-H"
                     /pseudo
                     /number=2
     CDS             <1..>259
                     /gene="HLA-H"
                     /note="MHC class I antigen HLA-H"
                     /codon_start=1
                     /pseudo
     gene            <1..>259
                     /gene="HLA-H"
                     /pseudo
BASE COUNT     52 a     83 c     85 g     37 t     2 others
ORIGIN

alignment_scores:
        Quality:   44.00              Length:   10
        Ratio:      4.400               Gaps:    0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x AF022160

Align seg 1/1 to: AF022160 from: 1 to: 259

    1 ArgGluAspArgIleAlaLeuArgTyr 10
      ||||||::||||||||||||||||||||
  223 CGAGAGAACCTGCGGATCGGCTCCGCTAC 252

seq_name: gb_pat:I14590

seq_documentation_block:
LOCUS       I14590       270 bp    DNA         PAT    26-SEP-1995
DEFINITION  Sequence 67 from patent US 5451512.
ACCESSION   I14590
VERSION     I14590.1   GI:997073
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 270)
AUTHORS     Apple,R.J., Bugawan,T.L. and Erlich,H.A.
TITLE       Methods and reagents for HLA class I A locus DNA typing
JOURNAL     Patent: US 5451512-A 67 19-SEP-1995;
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="unknown"
BASE COUNT     55 a     84 c     95 g     36 t
ORIGIN
```

```
ORIGIN

alignment_scores:
        Quality:   44.00              Length:   10
        Ratio:      4.400               Gaps:    0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x I14590

Align seg 1/1 to: I14590 from: 1 to: 270

    1 ArgGluAspArgIleAlaLeuArgTyr 10
      ||||||::||||||||||||||||||||
  222 CGAGAGAACCTGCGGATCGGCTCCGCTAC 251

seq_name: gb_pat:I14591

seq_documentation_block:
LOCUS       I14591       270 bp    DNA         PAT    26-SEP-1995
DEFINITION  Sequence 68 from patent US 5451512.
ACCESSION   I14591
VERSION     I14591.1   GI:997074
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 270)
AUTHORS     Apple,R.J., Bugawan,T.L. and Erlich,H.A.
TITLE       Methods and reagents for HLA class I A locus DNA typing
JOURNAL     Patent: US 5451512-A 68 19-SEP-1995;
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="unknown"
BASE COUNT     55 a     84 c     95 g     36 t
ORIGIN

alignment_scores:
        Quality:   44.00              Length:   10
        Ratio:      4.400               Gaps:    0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x I14591

Align seg 1/1 to: I14591 from: 1 to: 270

    1 ArgGluAspArgIleAlaLeuArgTyr 10
      ||||||::||||||||||||||||||||
  222 CGAGAGAACCTGCGGATCGGCTCCGCTAC 251

seq_name: gb_pat:I14592

seq_documentation_block:
LOCUS       I14592       270 bp    DNA         PAT    26-SEP-1995
DEFINITION  Sequence 69 from patent US 5451512.
ACCESSION   I14592
VERSION     I14592.1   GI:997075
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 270)
AUTHORS     Apple,R.J., Bugawan,T.L. and Erlich,H.A.
TITLE       Methods and reagents for HLA class I A locus DNA typing
JOURNAL     Patent: US 5451512-A 69 19-SEP-1995;
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="unknown"
BASE COUNT     55 a     84 c     95 g     36 t
ORIGIN
```

alignment_scores:
           Quality: 44.00   Length: 10
             Ratio: 4.400   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x I14592

Align seg 1/1 to: I14592 from: 1 to: 270

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      !!!!!:::!!!!!!!!!!!!!!!!!!!!!
  222 CGAGAGAACCTGCGGATCGGCGCTCCGCTAC 251

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-12 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000 1:27 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US08653294 -NCPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-08-653-294-12
Query length: 10
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 590.520000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
N_Geneseq_36:Q29167	+	44.00	162.25	0.1947	270	! HLA-Bw 52 exon 2 alpha-1 domain
N_Geneseq_36:Q01834	+	44.00	150.12	0.9226	1086	! Sequence encoding HLA-B51 anti
N_Geneseq_36:Q01822	+	44.00	150.12	0.9226	1086	! Sequence encoding HLA-Bw52 ant
N_Geneseq_36:Q05693	+	44.00	150.10	0.9255	1089	! HLA-B51 gene for production of
N_Geneseq_36:Q05701	+	44.00	150.10	0.9255	1089	! HLA-Bw52 gene for production o
N_Geneseq_36:Q12114	+	44.00	150.10	0.9255	1089	! HLA-Bw53 exon. HLA-Bw53 gene,
N_Geneseq_36:N70935	-	39.00	131.63	9.89	1026	! Sequence encoding the human hi
N_Geneseq_36:N70225	-	39.00	120.05	43.68	3874	! Sequence of genomic DNA encodi
N_Geneseq_36:T61639	-	39.00	115.47	78.59	6553	! HLA B27 consensus sequence. De
N_Geneseq_36:X51732	-	35.00	116.85	65.82	978	! DNA encoding a human secreted
N_Geneseq_36:X33945	-	35.00	116.38	151.02	2056	! Human HCMV inducible gene, SEQ
N_Geneseq_36:Q03811	-	34.00	126.17	19.91	217	! Turkey herpes virus (HVT) homol
N_Geneseq_36:T79497	-	34.00	112.63	113.03	1026	! Meripilus giganteus galactanas
N_Geneseq_36:X20647	-	34.00	111.68	127.78	1145	! Polynucleotide sequence from t
N_Geneseq_36:V35645	+	34.00	105.28	290.30	2386	! Cladosporium oxysporum glucose
N_Geneseq_36:X13670	-	33.00	108.49	192.22	1067	! Enterococcus faecalis genome c
N_Geneseq_36:X34093	+	33.00	105.41	285.47	1520	! Mycobacterium species nucleic
N_Geneseq_36:T30773	-	33.00	102.03	440.33	2240	! CD40 associated protein (CAP)-
N_Geneseq_36:T31273	-	33.00	101.58	466.56	2359	! LMP1 associated protein LAP1 g
N_Geneseq_36:T90123	-	33.00	99.73	591.74	2918	! Human CRAF1 (TRAF-3) cDNA. Pro
N_Geneseq_36:Q24977	-	33.00	94.79	1.1e+03	5140	! DNA encoding soluble mannose r
N_Geneseq_36:V30458_0	+	33.00	68.10	3.3e+04	110000	! Continuation (8 of 17) of
N_Geneseq_36:V30459_0	+	33.00	68.10	3.3e+04	110000	! Continuation (14 of 17) of
N_Geneseq_36:X30161	-	32.00	105.24	291.65	1002	! Human secreted protein gene 17
N_Geneseq_36:V84507	-	32.00	103.92	345.82	1167	! Human secreted protein gene 97
N_Geneseq_36:X07101	+	32.00	102.84	396.88	1320	! Staphylococcus aureus mutant F
N_Geneseq_36:Q97686	-	32.00	85.41	3.7e+03	9757	! Infectious rubella virus RNA.
N_Geneseq_36:T89642	-	32.00	85.41	3.7e+03	9759	! Infectious rubella virus cDNA
N_Geneseq_36:V34766	-	32.00	85.41	3.7e+03	9759	! Rubella virus RA27/3 genomic s
N_Geneseq_36:V74414	-	32.00	82.23	5.6e+03	14051	! Staphylococcus aureus contig
N_Geneseq_36:V21209_07	+	32.00	64.30	5.3e+04	110000	! Continuation (8 of 17) of
N_Geneseq_36:V21209_13	-	32.00	64.30	5.3e+04	110000	! Continuation (14 of 17) of
N_Geneseq_36:Q76405	-	31.00	114.94	84.10	213	! Human genome fragment. (Preferi
N_Geneseq_36:T09204	-	31.00	112.04	121.95	297	! Virulence factor sequence taken
N_Geneseq_36:N50356	+	31.00	112.89	394.47	849	! Sequence encoding fused antibod
N_Geneseq_36:V37348	+	31.00	102.11	435.71	928	! Streptococcus pneumoniae coding
N_Geneseq_36:T05843	-	31.00	102.06	438.86	934	! Repeat sequence of cps gene loc
N_Geneseq_36:N80055	+	31.00	101.37	479.49	1011	! Beta-glucuronidase-contg inser
N_Geneseq_36:Q05916	+	31.00	101.37	479.49	1011	! Promoter sequence and N-termin
N_Geneseq_36:T01865	+	31.00	100.52	534.40	1114	! Fc(epsilon) CH2'-CH4 coding se
N_Geneseq_36:V18352	+	31.00	100.16	559.67	1161	! Human V28 seven transmembrane

N_Geneseq_36:Q66170	+	31.00	100.15	560.20	1162	! Seven transmembrane recepto
N_Geneseq_36:V21128	+	31.00	99.28	626.34	1284	! Human C epsilon exon. New 1
N_Geneseq_36:Q87474	+	31.00	99.12	639.44	1308	! Human IgE Fc chain (amino a
N_Geneseq_36:Q91170	+	31.00	99.12	639.44	1308	! Human IgE Fc chain (amino a

seq_name: N_Geneseq_36:Q29167

seq_documentation_block:
ID Q29167 standard; DNA; 270 BP.
AC Q29167;
DT 09-MAR-1993 (first entry)
DE HLA-Bw 52 exon 2 alpha-1 domain.
KW Human leukocyte antigen; transgenic; germ cells; somatic cells;
KW expression; ss.
PN J04091731-A.
PD 25-MAR-1992.
PF 03-AUG-1990; 207329.
PR 03-AUG-1990; JP-207329.
PA (OLYU) OLYMPUS OPTICAL CO.
DR WPI; 92-342893/42.
PT Transgenic non-human mammalian HLA-Bw 52 gene - useful for
PT analysis of expression of gene structure, and prodn. of
PT mouse model of human disease
PS Disclosure: Fig 1; 8pp; Japanese.
CC The sequence shows the exon 2 alpha-1-domain of the human leukocyte
CC antigen-Bw 52 gene. The complete gene may be introduced into non-
CC human mammals, pref. rat or mouse, or their ancestors at the primary
CC developmental biological step via transplantation into the zygote or
CC embryo to generate transgenic non-human mammals incorporating the
CC HLA-Bw 52 gene in both their germ cells and somatic cells. Transgenic
CC non-human mammals contg. HLA-Bw 52 are useful for the analysis of
CC expression of the gene, its structure, and prodn. of mouse models of
CC human disease.   See also Q29166-72.
SQ Sequence 270 BP;   59 A;   88 C;   86 G;   37 T;

alignment_scores:
        Quality: 44.00   Length: 10
        Ratio: 4.400   Gaps: 0
  Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q29167

Align seg 1/1 to: Q29167 from: 1 to: 270

        1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
          |||||:::||||||||||||||||||||||||
      222 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 251

seq_name: N_Geneseq_36:Q01834

seq_documentation_block:
ID Q01834 standard; DNA; 1086 BP.
AC Q01834;
DT 19-MAR-1991 (first entry)
DE Sequence encoding HLA-B51 antigen.
KW Probe; HLA class I DNA; immunogen; ss.
OS Homo sapiens.
PN EP354580-A.
PD 14-FEB-1990.
PF 10-AUG-1989.
PR 11-AUG-1988; JP-200758.
PA (OLYU) Olympus Optical Co., Ltd.
PI Kano K, Takiguchi;
DR WPI; 90-046289/07.
PT New DNA for class 1 human leucocyte antigens and derived probes and
PT transformed cells, useful for DNA typing, as immunogens etc.
PS Claim 1; Page 11; 23pp; English.
CC The HLA class I DNA can be used a source of probes for use in DNA
CC typing. Transformed cells, which are useful as immmunogens, can be
CC obtained by introducing these DNAs into eucaryotic cells.
SQ Sequence 1086 BP;   224 A;   334 C;   356 G;   172 T;

```
alignment_scores:
    Quality:    44.00              Length:          10
    Ratio:      4.400              Gaps:             0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q01834

Align seg 1/1  to: Q01834  from: 1  to: 1086

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      |||||::::|||||||||||||||||
  294 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 323

seq_name: N_Geneseq_36:Q01822

seq_documentation_block:
ID  Q01822 standard; DNA; 1086 BP.
AC  Q01822;
DT  19-MAY-1991 (first entry)
DE  Sequence encoding HLA-Bw52 antigen.
KW  Probe; HLA class I DNA; immunogen; ss.
OS  Homo sapiens.
FH  Key          Location/Qualifiers
FT  cds          1..1086
FT               /tag= a
PN  EP-354580-A.
PD  14-FEB-1990.
PF  10-AUG-1989.
PR  11-AUG-1988; JP-200758.
PA  (OLYU) Olympus Optical Co., Ltd.
PI  Kano K, Takiguchi;
WPI: 90-046288/07.
DR  P-PSDB; R03142.
PT  New DNA for class 1 human leucocyte antigens and derived probes and
PT  transformed cells, useful for DNA typing, as immunogens etc.
PS  Claim 2; ppl-12; 23pp; English.
CC  The HLA class I DNA can be used a source of probes for use in DNA
CC  typing. Transformed cells, which are useful as immunogens, can be
CC  obtained by introducing these DNAs into eucaryotic cells.
SQ  Sequence 1086 BP;   223 A;   335 C;   358 G;   170 T;


alignment_scores:
    Quality:    44.00              Length:          10
    Ratio:      4.400              Gaps:             0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q01822

Align seg 1/1  to: Q01822  from: 1  to: 1086

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      |||||::::|||||||||||||||||
  294 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 323

seq_name: N_Geneseq_36:Q05693

seq_documentation_block:
ID  Q05693 standard; DNA; 1089 BP.
AC  Q05693;
DT  03-JAN-1991 (first entry)
DE  HLA-Bw52 gene for production of monoclonal antibodies.
KW  Alloptype specific monoclonal anti-HLA antibodies,
KW  transgenic animals; HLA-B51 gene; ss.
FH  Key          Location/Qualifiers
FT  exon         1..73
FT               /tag= a
FT               /number=1
FT  exon         74..343
```

```
FT                             /tag= b
FT                             /number=2
FT                             /note="alpha 1-domain"
FT  exon         344..619
FT                             /tag= c
FT                             /number=3
FT                             /note="alpha 2-domain"
FT  intron       620..895
FT                             /tag= d
FT                             /number=4
FT                             /note="alpha 3-domain"
FT  exon         896..1012
FT                             /tag= e
FT                             /number=5
FT  exon         1013..1042
FT                             /tag= f
FT                             /number=6
FT  exon         1043..1089
FT                             /tag= g
FT                             /number=7
PN  EP-383183-A.
PD  22-AUG-1990.
PF  07-FEB-1990; 102424.
PR  08-FEB-1989; JP-029313.
PA  (OLYU ) OLYMPUS OPTICAL KK.
PI  Takiguchi M;
DR  WPI: 90-255479/34.
PT  Allotype specific monoclonal anti- HLA antibodies prodn. - using
PT  hybridomas derived from transgenic animals carrying HLA gene and
PT  immunised with HLA antigen of different allotype
PS  Disclosure; Fig 1 A-G; 20pp; English.
CC  The human HLA-B51 gene was injected into fertilised mouse eggs and
CC  then these introduced into the uterus of a pseudo pregnant mouse.
CC  The young were tested to ensure incorporation of the gene into the
CC  chromosome, and one of them mated 3 times with a normal male to
CC  produece 16 young, seven of which carried the HLA-B51 gene.
CC  The transgenic offspring were immunised with HLA antigen.
CC  The spleen lymphocytes were fused with myeloma cells. Hybridomas
CC  producing antibodies were selected.
CC  See also Q05701.
SQ  Sequence 1089 BP;   224 A;   335 C;   357 G;   173 T;


alignment_scores:
    Quality:    44.00              Length:          10
    Ratio:      4.400              Gaps:             0
Percent Similarity: 100.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q05693

Align seg 1/1  to: Q05693  from: 1  to: 1089

    1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
      |||||::::||||||||||||||||||
  295 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 324

seq_name: N_Geneseq_36:Q05701

seq_documentation_block:
ID  Q05701 standard; DNA; 1089 BP.
AC  Q05701;
DT  03-JAN-1991 (first entry)
DE  HLA-Bw52 gene for production of monoclonal antibodies.
KW  Alloptype specific monoclonal anti-HLA antibodies; hybridomas;
KW  transgenic animals; HLA-Bw52 gene; ss.
FH  Key          Location/Qualifiers
FT  exon         1..73
FT               /tag= a
FT               /number=1
FT  exon         74..343
FT               /tag= b
FT               /number=2
```

```
FT   exon            /note="alpha 1-domain"
FT                   344..619
FT                   /*tag= c
FT                   /number=3
FT   exon            /note="alpha 2-domain"
FT                   620..895
FT                   /*tag= d
FT                   /number=4
FT   exon            /note="alpha 3-domain"
FT                   896..1012
FT                   /*tag= e
FT                   /number=5
FT   exon            1013..1042
FT                   /*tag= f
FT                   /number=6
FT   exon            1043..1089
FT                   /*tag= g
FT                   /number=7
PN   EP-383183-A.
PD   22-AUG-1990.
PF   07-FEB-1990; 102424.
PR   08-FEB-1989; JP-029313.
PA   (OLYU ) OLYMPUS OPTICAL KK.
PI   Takiguchi M;
DR   WPI: 90-255479/34.
PT   Allotype specific monoclonal anti- HLA antibodies prodn. - using
PT   hybridomas derived from transgenic animals carrying HLA gene and
PT   immunised with HLA antigen of different allotype
PS   Disclosure; Fig 1 A'G; 20pp; English.
CC   The human HLA-Bw52 gene was introduced into mouse L cells and
CC   then these cells used to immunise one of the transgenic mice
CC   (See Q05693).
CC   The spleen lymphocytes were fused with myeloma cells (P3x63-Ag8.653).
CC   Hybridomas producing antibodies were selected.
SQ   Sequence 1089 BP;    336 C;    223 A;    359 G;    171 T;

alignment_scores:
       Quality:      44.00          Length:       10
        Ratio:        4.400           Gaps:        0
Percent Similarity: 100.000     Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q05701

Align seg 1/1 to: Q05701 from: 1 to: 1089

  1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
    ||||||::|||||||||||||||||||
295 CGAGAGAACCTGCGGATCGCGCTCCCGCTAC 324

seq_name: N_Geneseq_36:Q12114.

seq_documentation_block:
ID   Q12114 standard; DNA; 1089 BP.
AC   Q12114;
DT   29-AUG-1991 (first entry)
DE   HLA-Bw53 exon.
KW   Human leukocyte antigen; probe; major histocompatibility complex;
KW   MHC; class I: ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1..1089
FT                   /*tag= a
PN   J03112487-A.
PD   14-MAY-1991.
PF   22-SEP-1989; 247697.
PR   22-SEP-1989; JP-247697.
PA   (OLYU ) OLYMPUS OPTICAL KK.
     WPI: 91-182991/25.
DR   P-PSDB: R12463.
PT   HLA-Bw53 gene, DNA probe and transformant cells - used for
PT   immunisation, identifying specificity of antiserum etc.
```

```
PS   Claim 1; Page 1; 11pp; Japanese.
CC   Probes comprising part of the sequence can be used to identify
CC   Class I genes. The DNA can be expressed for immunisation of
CC   animals and prodn. of monoclonal antibodies specific for the
CC   HLA-Bw3 antigen. See also J03112485 and J03112486.
SQ   Sequence 1089 BP;    222 A;    337 C;    356 G;    174 T;

alignment_scores:
       Quality:      44.00          Length:       10
        Ratio:        4.400           Gaps:        0
Percent Similarity: 100.000     Percent Identity: 90.000

alignment_block:
US-08-653-294-12 x Q12114

Align seg 1/1 to: Q12114 from: 1 to: 1089

  1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
    ||||||::|||||||||||||||||||
295 CGAGAGAACCTGCGGATCGCGCTCCCGCTAC 324

seq_name: N_Geneseq_36:N70935

seq_documentation_block:
ID   N70935 standard; DNA; 1026 BP.
AC   N70935;
DT   10-APR-1991 (first entry)
DE   Sequence encoding the human histocompatibility antigen HLA B27.
KW   Rheumatic disorder; genetic screening; diagnosis;
KW   ankylosing spondylitis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .1026
PN   DE3542024-A.
PD   04-JUN-1987.
PF   28-NOV-1985; 542024.
PR   28-NOV-1985; DE-542024.
PR   21-DEC-1985; DE-545576.
PA   (BEHW ) BEHRINGWERKE AG.
PI   Riethmuller G, Meo T, Weiss E, Szots H;
DR   WPI: 87-157893/23.
DR   P-PSDB; P70590.
PT   DNA coding for antigen HLA B27 - and diagnostic reagents contg.
PT   such DNA, antigen or antibody
PS   Claim 2; Page 4: 5pp; German.
CC   The DNA may be used as a hybridisation probe for detecting the HLA
CC   B27 gene, e.g. for assessing susceptibility to rheumatic disorders
CC   such as ankylosis spondylitis, or may be used to transform cells
CC   for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27
CC   antibody in human serum, or to produce mono- or polyclonal HLA B27
CC   antibodies for use in immunoassay.
SQ   Sequence 1026 BP;    213 A;    307 C;    344 G;    162 T;

alignment_scores:
       Quality:      39.00          Length:       10
        Ratio:        4.875           Gaps:        0
Percent Similarity: 80.000     Percent Identity: 80.000

alignment_block:
US-08-653-294-12 x N70935

Align seg 1/1 to: N70935 from: 1 to: 1026

  1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
    ||||||| ||||||
223 CGAGAGGACCTGCGGACCCTGCTCCCGCTAC 252

seq_name: N_Geneseq_36:N70225

seq_documentation_block:
ID   N70225 standard; DNA; 3874 BP.
```

```
AC   N70225;
DT   03-APR-1991 (first entry)
DE   Sequence of genomic DNA encoding human histocompatibility antigen
DE   HLA-B 27.
KW   Ankylosing spondylitis; rheumatic disorder; diagnosis; ss.
OS   Homo sapiens.
FH   Key         Location/Qualifiers
FT   intron      518..590
FT               /*tag= a
FT   intron      720..989
FT               /*tag= b
FT   intron      1090..1506
FT               /*tag= c
FT   intron      1932..2357
FT               /*tag= d
FT   intron      2450..2566
FT               /*tag= e
FT   intron      3009..3041
FT               /*tag= f
FT   intron      3148..3191
FT               /*tag= g
PN   EP-226069-A.
PD   24-JUN-1987.
PF   21-NOV-1986; 116139.
PR   01-JAN-1985; DE-542024.
PR   21-DEC-1985; DE-545576.
PA   (BEHW ) BEHRINGWERKE AG.
PI   Szots H, Weiss E, Dorner C, Lang M, Meo T, Riethmuller G;
WPI  87-171469/25.
DR   P-PSDB; P70155.
PT   DNA coding for human histocompatibility antigen HLA-B 27 - useful
PT   for diagnosis and antigen and antibody prodn.
PS   Claim 1; p6; 13pp; German.
CC   THe DNA may be used to detect the HLA-B 27 gene (opt. mutated) in
CC   human genetic material. The HLA-B 27 may be used to detect anti-HLA-
CC   B 27 antibodies in human serum. The antibodies may be used to
CC   determine HLA-B 27 levels in human serum, e.g. for diagnosis of
CC   rheumatic disorders, esp. ankylosing spondylitis.
SQ   Sequence 3874 BP; 751 A; 1094 C; 1171 G; 858 T;

alignment_scores:
         Quality:    39.00              Length:       10
         Ratio:       4.875             Gaps:          0
Percent Similarity:  80.000      Percent Identity: 80.000

alignment_block:
US-08-653-294-12 x N70225     ..

Align seg 1/1  to: N70225  from: 1  to: 3874

     1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
       ||||||||||   ||||||||
   941 CGAGAGGACCTGCGGACCCTGCTCCCGCTAC 970

seq_name: N_Geneseq_36:T61639   ..

seq_documentation_block:
ID   T61639 standard; DNA; 6553 BP.
AC   T61639;
DT   05-JUN-1997 (first entry)
DE   HLA B27 consensus sequence.
KW   HLA B27; seronegative spondylarthropathy; ankylosing spondylitis;
KW   Reiter's syndrome; arthritis; acute anterior uveitis; diagnosis;
KW   ss; ds.
OS   Homo sapiens.
FH   Key         Location/Qualifiers
FT   mrna        3968..6653
FT               /*tag= a
FT               /note= "HLA-B27 3' flanking region, downstream of
FT               3' untranslated region"
FT   mrna        4112..4556
FT               /*tag= b
```

```
FT               /note= "3' flanking region diagnostic for genetic
FT               predisposition to SNSA"
FT   mrna        4270..4556
FT               /*tag= b
FT               /note= "3' flanking region diagnostic for genetic
FT               predisposition to SNSA"
FT   misc_difference 4495
FT               /*tag= d
FT               /note= "absence of cytosine at this site is
FT               indicative of a predisposition to SNSA"
PN   WO9709450-A1.
PD   13-MAR-1997.
PF   16-AUG-1996; U13256.
PR   01-SEP-1995; US-522942.
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
PI   Tyan DB;
DR   WPI; 97-192924/17.
PT   Detecting pre-disposition to seronegative spondylarthropathies -
PT   from the absence of a C residue at a specific position in the
PT   3'-flanking region of the HLA B27 allele
PS   Claim 1; Page 52-56; 68pp; English.
CC   Genetic predisposition to seronegative spondyloarthropathies (SNSA)
CC   is detected by determining the absence of a cytosine nucleotide in
CC   the 3' flanking region (see also T61647-48) of an HLA-B gene at a
CC   position corresponding to nucleotide 4495 of the HLA-B27 consensus
CC   sequence given in T61639.  Probes and primers (see also T61640-46)
CC   based on this region can be used in diagnostic assays to detect the
CC   genetic predisposition to SNSA, and permit the distinction of B27+
CC   individuals who are resistant to SNSA from B27+ normal individuals
CC   who are susceptible (but as yet unaffected) to such diseases.
SQ   Sequence 6553 BP; 1443 A; 1619 C; 2017 G; 1474 T;

alignment_scores:
         Quality:    39.00              Length:       10
         Ratio:       4.875             Gaps:          0
Percent Similarity:  80.000      Percent Identity: 80.000

alignment_block:
US-08-653-294-12 x T61639     ..

Align seg 1/1  to: T61639  from: 1  to: 6553

     1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
       ||||||||||   ||||||||
  1102 CGAGAGGACCTGCGGACCCTGCTCCCGCTAC 1131

seq_name: N_Geneseq_36:X51732   ..

seq_documentation_block:
ID   X51732 standard; DNA; 978 BP.
AC   X51732;
DT   17-JUN-1999 (first entry)
DE   DNA encoding a human secreted protein.
KW   Human secreted protein; cancer; immune disorder; infection;
KW   inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW   restenosis; autoimmune disorder; Alzheimer's disease;
KW   peripheral neuropathy; trauma; spinal cord injury; allergy;
KW   hematopoietic disorder; skeletal disorder; neurological disorder;
KW   arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW   transplant rejection; ss.
OS   Homo sapiens.
PN   WO911293-A1.
PD   11-MAR-1999.
PF   03-SEP-1998; U18360.
PR   12-SEP-1997; US-058974.
PR   12-SEP-1997; US-057626.
PR   05-SEP-1997; US-057663.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-058666.
PR   12-SEP-1997; US-058667.
PR   12-SEP-1997; US-058973.
PA   (HUMA-) HUMAN GENOME SCI INC.
```

PI   Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS;
PI   Rosen GA, Ruben SM, Shi Y;
DR   WPI; 99-204988/17.
DR   P-PSDB; Y12945.
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. neurological disorders,
PT   tumours, immune disorders, inflammation or haematological disorders
PS   Claim 1; Page 170; 215pp; English.
CC   X51701-55 encode human secreted proteins. The polynucleotides and
CC   their corresponding secreted polypeptides are useful for preventing,
CC   treating or ameliorating medical conditions, e.g. by protein or gene
CC   therapy. Pathological conditions can also be diagnosed by determining
CC   the amount of the new polypeptides in a sample or by determining the
CC   presence of mutations in the new polynucleotides. Specific uses are
CC   described for each polynucleotide, based on which tissues they are
CC   most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, immune disorders, infection,
CC   inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC   restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC   neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC   disorders, skeletal disorders, neurological disorders, arthritic
CC   disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC   rejection. The polypeptides are also useful for identifying their
CC   binding partners.
SQ   Sequence 978 BP; 285 A; 256 C; 197 G; 240 T;

alignment_scores:
        Quality: 35.00   Length: 10
        Ratio: 3.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-12 x X51732/rev ..

Align seg 1/1 to reverse of: X51732 from: 1 to: 978

   1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
     ||||||::||||||||| ||||||::
  99 AGGGAGGAGCTGAGAATCCAACTGCCGTGG 70

seq_name: N_Geneseq_36:X33945

seq_documentation_block:
ID   X33945 standard; DNA; 2056 BP.
AC   X33945;
DT   30-JUN-1999 (first entry)
DE   Human HCMV inducible gene, SEQ ID NO 17.
KW   HCMV inducible gene; cig; human; human cytomegalovirus; interferon;
KW   anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
KW   drug screening; ds.
OS   Homo sapiens.
PN   WO9913075-A2.
PD   18-MAR-1999.
PF   08-SEP-1998; U18638.
PR   22-SEP-1997; US-059725.
PR   08-SEP-1997; US-058180.
PA   (UYPR-) UNIV PRINCETON.
PI   Cong J, Schenk T, Zhu H;
DR   WPI; 99-243729/20.
DR   P-PSDB; Y05375.
PT   New isolated human genes
PS   Claim 2; Page 126-129; 184pp; English.
CC   This sequence represents a human gene of the invention, that is induced
CC   to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC   genes (cig or cigs). The invention also relates to genes that are
CC   repressed in the presence of HCMV infection, designated HCMV-repressible
CC   genes (crg or crgs). The products can be used to obtain agents which can
CC   be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC   also be used for the development of drugs that would allow for higher
CC   dosage IFN treatments without the concomitant toxicity normally
CC   associated with administering high levels of IFN. The products can also
CC   be used for detection, diagnosis and drug screening.

SQ   Sequence 2056 BP; 645 A; 437 C; 526 G; 448 T;

alignment_scores:
        Quality: 35.00   Length: 10
        Ratio: 3.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-12 x X33945 ..

Align seg 1/1 to: X33945 from: 1 to: 2056

   1 ArgGluAspLeuArgIleAlaLeuArgTyr 10
     ||||| |||::::||||||||||||
1665 AGGGAAGCTTTGCATGTTGCTCTCTAAGGTAC 1694

seq_name: N_Geneseq_36:Q03811

seq_documentation_block:
ID   Q03811 standard; DNA; 217 BP.
AC   Q03811;
DT   28-AUG-1990 (first entry)
DE   Turkey herpes virus (HVT) homologue of the VZV62/HSV-1 IE 175 gene.
KW   Infectious laryngotracheitis virus; ILTV; Marek's disease virus; MDV;
KW   Herpes virus of turkeys; HVT; vaccine; ribonucleotide reductase; ds.
OS   Turkey herpes virus.
PN   WO9002802-A.
PD   22-MAR-1990.
PF   13-SEP-1989; 001075.
PR   13-SEP-1988; GB-021441.
PA   (ANIM-) Inst Animal Health.
PI   Griffin AM, Ross LJN, Scott SD, Binns NM;
DR   WPI; 90-115990/15.
DR   P-PSDB; R05832.
PT   Non-essential viral nucleotide sequences - identified in Herpes virus of turkeys, Marek's disease virus and
PT   infectious laryngotracheitis virus.
PS   Disclosure; Fig 11; 111pp; English.
CC   The gene products may be used as antigens to provide in part a
CC   multivalent vaccine, protecting against ILTV, MDV and HVT.
CC   See also WO9002803-A.
SQ   Sequence 217 BP; 47 A; 59 C; 62 G; 48 T; 1 Others;

alignment_scores:
        Quality: 34.00   Length: 8
        Ratio: 4.250   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
US-08-653-294-12 x Q03811 ..

Align seg 1/1 to: Q03811 from: 1 to: 217

   3 AspLeuArgIleAlaLeuArgTyr 10
     ||||||||||:::||||:||||||
150 GACCTCCGAGTGGCATTCCGCTAC 173

seq_name: N_Geneseq_36:T79497

seq_documentation_block:
ID   T79497 standard; cDNA; 1026 BP.
AC   T79497;
DT   26-JAN-1998 (first entry)
DE   Meriplus giganteus galactanase cDNA.
KW   Meriplus giganteus; galactanase; EC 3.2.1.89; arabinogalactanase;
KW   endo-1,4-beta-galactanase; transformant; viscosity; wine production;
KW   arabinogalactan endo-1,4-beta-galactosidase; fruit juice;
KW   vegetable juice; animal feed; depectinisation; ds.
OS   Meriplus giganteus.
FH   Key   Location/Qualifiers
FT   sig_peptide   1..54

FT   mat_peptide   /*tag= a
FT   55..1026
FT   /*tag= b
FT   /product= galactanase
FT   /EC_number= 3.2.1.89
PN   WO9732013-A1.
PD   04-SEP-1997.
PF   28-FEB-1997; DK0091.
PR   01-MAR-1996; DK-000234.
PA   (NOVO) NOVO-NORDISK AS.
PI   Andersen LN, Clausen IG, Kauppinen MS, Kofod LV;
DR   WPI: 97-448685/41.
DR   P-PSDB: W23140.
PT   DNA sequence encoding fungal galactanase - useful in production of
PT   wine or modification of animal feed, e.g. depectinisation, and
PT   reducing viscosity of plant cell wall derived material
PS   Claim 1: Pages 34-36; 49pp; English.
CC   This cDNA sequence encodes a novel galactanase isolated from a
CC   M. giganteus library constructed in E. coli. Positive cDNA inserts were
CC   identified on SC-agar plates using the AZCL xylan assay. cDNA inserts
CC   were amplified directly from yeast colonies. In order to express the
CC   galactanase in Aspergillus, the DNA was digested with appropriate
CC   restriction enzymes, size fractionated on a gel and a fragment
CC   corresponding to the galactanase gene purified. The gene was
CC   subsequently ligated into pHD414 and digested with appropriate
CC   restriction enzymes to give pA2G55. After amplification of the DNA in
CC   E. coli the plasmid was transformed into Aspergillus oryzae and
CC   transformant activity was analysed. The galactanase can be used to
CC   reduce the viscosity of plant cell wall derived material, thus having
CC   implications in wine production, the preparation of fruit or vegetable
CC   juice or for the modification of animal feed allowing a significant
CC   improvement in the in vivo breakdown of plant cell wall material e.g.
CC   depectinisation.
SQ   Sequence 1026 BP;   214 A;   331 C;   285 G;   196 T;

alignment_scores:
        Quality: 34.00   Length: 9
        Ratio: 4.250   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-08-653-294-12 x T79497/rev ..
Align seg 1/1 to reverse of: T79497 from: 1 to: 1026

            2 GluAspLeuArgIleAlaLeuArgTyr 10
              ||||||||||||::||| |||::::
          564 GAGGATCTTCGGGTTGCCTGCCGCTTT 538

seq_name: N_Geneseq_36:X20647

seq_documentation_block:
ID   X20647 standard; DNA; 1145 BP.
AC   X20647;
DT   05-MAY-1999 (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum. .
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI: 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1: Page 799-800; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,

CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
SQ   Sequence 1145 BP;   242 A;   267 C;   374 G;   257 T;

alignment_scores:
        Quality: 34.00   Length: 9
        Ratio: 4.250   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 77.778

alignment_block:
US-08-653-294-12 x X20647/rev ..
Align seg 1/1 to reverse of: X20647 from: 1 to: 1145

            2 GluAspLeuArgIleAlaLeuArgTyr 10
              :::||||||||||||||||| |||
          839 AAAGATCTTAGGATAGCTCTTGCGTAC 813

seq_name: N_Geneseq_36:V35645

seq_documentation_block:
ID   V35645 standard; DNA; 2386 BP.
AC   V35645;
DT   08-SEP-1998 (first entry)
DE   Cladosporium oxysporum glucose oxidase encoding DNA.
KW   Glucose oxidase; Cladosporium oxysporum; enzyme; bread improver;
KW   antimicrobial agent; toothpaste; detergent; stain removal;
KW   dough additive; hydrogen peroxide generator; ss.
OS   Cladosporium oxysporum.
FH   Key   Location/Qualifiers
FT   5'UTR   1..350
FT   /*tag= a
FT   CDS   351..2192
FT   /*tag= b
FT   /product= "glucose oxidase"
FT   sig_peptide   351..419
FT   /*tag= c
FT   /note= "predicted signal sequence"
FT   mat_peptide   420..2189
FT   /*tag= d
FT   3'UTR   2190..2386
FT   /*tag= e
PN   WO9820136-A1.
PD   14-MAY-1998.
PF   03-NOV-1997; U20174.
PR   07-NOV-1996; US-746257.
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO-NORDISK AS.
PI   Berka RM, Cherry JR, Halkier T;
DR   WPI: 98-286952/25.
DR   P-PSDB: W60593.
PT   New nucleic acid encoding glucose oxidase active at acidic pH, from
PT   Cladosporidium - and related vectors and host cells, producing
PT   enzyme useful as bread improver, antimicrobial additive for
PT   toothpaste, etc. and hydrogen peroxide generator in detergents
PS   Claim 4; Fig 2A-B; 83pp; English.
CC   This DNA encodes a Cladosporium oxysporum glucose oxidase. Host cells
CC   containing a construct comprising the glucose oxidase encoding nucleic
CC   acid sequence are used to produce recombinant
CC   glucose oxidase. Preferred fragments of the nucleic acid are present in
CC   pGOX4A and pGOX6A, contained in E. coli deposited as NRRL B-21628 and
CC   B-21629. The host cells are particularly Fusarium. The recombinant
CC   glucose oxidase is useful as dough additive to improve the gluten
CC   quality. It can be used as additive for toothpaste (particularly used
CC   with a thiocyanate and lactoperoxidase to generate antimicrobial
CC   oxythiocyanate anion), mouthwash, denture cleaners, soaps, hair and body
CC   care products. It can be added to cleaning solutions for contact lenses
CC   as bleach or as an antibacterial agent. It is also useful as a hydrogen
CC   peroxide generator in laundry and dishwashing detergents, particularly
CC   for stain removal.

SQ   Sequence   2386 BP;   583 A;   707 C;   581 G;   515 T;

alignment_scores:
        Quality: 34.00   Length: 9
          Ratio: 4.250   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 77.778

alignment_block:
US-08-653-294-12 x V35645   ..

Align seg 1/1 to: V35645   from: 1   to: 2386

      1 ArgGluAspLeuArgIleAlaLeuArg 9
        ||||||||||||||| .::||||||
    605 CGTGAGGACTTACGGACAAGCCTTCGA 631

OM of: US-08-653-294-10 to: EST:*   out_format : pfs
Date: Feb 8, 2000 4:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -O=/cgul_1/USPTO.spool/US08653294/runat_04022000_160700_15770/app-query.fasta.1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cd1
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-08-653-294-10
 Query length: 10
 Database: EST:*
 Database sequences: 4538634
 Database length: 1887881982
 Search time (sec): 8553.360000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_est8:C03945	+	44.00	163.32	2.48	232	! C03945 C03945 Human heart cDNA
gb_est11:AA151891	+	44.00	162.59	2.73	255	! AA151891 zo01f06.r1 Stratagene
gb_est3:AA263158	+	44.00	161.78	3.03	283	! AA263158 PMY0534 KGI-a Lambda Z
gb_est6:D82221	+	44.00	159.59	4.01	375	! D82221 HUMHBC4626 Human pancrea
gb_est10:AA147151	+	44.00	156.18	6.21	581	! AA147151 zo32d06.r1 Stratagene
gb_est26:AI359260	+	44.00	155.70	6.60	618	! AI359260 qy27b07.x1 NCI_CGAP_Br
gb_est31:AI696864	+	44.00	154.22	7.99	748	! AI696864 wc74h11.x1 NCI_CGAP_Pa
gb_est13:AA332511	+	40.00	144.43	28.03	360	! AA332511 EST36483 Embryo, 8 wee
gb_est38:AL036690	+	39.00	146.35	21.91	171	! AL036690 DKFZp564D2453.r1 564 (
gb_gss9:AQ163964	+	39.00	140.75	44.89	351	! AQ163964 HS_2270_B1_A09_MF CIT
gb_est23:AI124815	+	39.00	139.47	52.93	414	! AI124815 am56e06.x1 Johnston fr
gb_est15:AQ614213	+	39.00	138.03	63.64	498	! AQ614213 HS_5123_B1_B06_SP6E RF
gb_est21:AA989542	-	38.00	135.83	84.43	402	! AA989542 am64d02.s1 Barstead sp
gb_est6:D82189	-	38.00	135.58	87.15	415	! D82189 HUMHBC4524 Human pancrea
gb_est38:AW092686	+	38.00	135.13	92.39	440	! AW092686 EST285866 tomato mixed
gb_gss13:AQ435812	+	38.00	135.00	93.85	447	! AQ435812 HS_5063_A1_B10_SP6E RF
gb_est21:AA975627	-	38.00	134.05	106.00	505	! AA975627 oq63b05.s1 NCI_CGAP_Ki
gb_gss9:AQ164696	-	38.00	133.80	109.56	522	! AQ164696 HS_3006_B2_D07_T7 CIT
gb_gss4:AQ721173	-	38.00	133.11	119.61	570	! AQ721173 HS_5564_A2_A04_T7A RPC
gb_est22:AI058482	+	37.00	139.32	53.94	156	! AI058482 UI-R-C1-kg-c-05-0-UI.s
gb_est21:AA952680	+	37.00	135.08	92.89	269	! AA952680 TENS1864 T. cruzi epin
gb_est13:AA366310	+	37.00	133.06	120.44	349	! AA366310 EST77241 Pancreas tumo
gb_est20:AA850199	+	37.00	132.93	122.50	355	! AA850199 EST192966 Normalized r
gb_est1:T29557	+	37.00	132.82	124.22	360	! T29557 EST84446 Human Colon mous
gb_est10:AA178827	-	37.00	132.63	127.32	369	! AA178827 mt68h02.r1 Soares mous
gb_est8:AA053595	-	37.00	132.08	136.62	396	! AA053595 zl73c04.r1 Stratagene
gb_est22:AU014732	-	37.00	131.98	138.34	401	! AU014732 AU014732 Mouse two-cel
gb_gss9:AA099810	-	37.00	131.94	139.03	403	! AA099810 zl81h06.r1 Stratagene
gb_est11:AA239196	+	37.00	131.35	150.04	435	! AA239196 mx89c04.r1 Soares mous
gb_est19:AA748746	+	37.00	131.29	151.07	438	! AA748746 ny06g04.s1 NCI_CGAP_GC
gb_est2:R13904	+	37.00	130.93	158.30	459	! R13904 yf62c03.r1 Soares infant
gb_est24:AI172015	+	37.00	129.79	183.06	531	! AI172015 EST218010 Normalized r
gb_gss12:AQ415562	+	37.00	129.61	187.54	544	! AQ415562 RPCI-11-204A14.TJ RPCI
gb_est22:AU014732	-	37.00	129.56	188.57	547	! AU014732 AU014732 Mouse two-cel
gb_est44:AW220783	+	37.00	129.56	188.57	547	! AW220783 EST297252 tomato fruit
gb_est25:AU045014	+	37.00	129.15	198.89	577	! AU045014 AU045014 Mouse sixteen
gb_est10:AA143780	+	37.00	128.65	211.95	615	! AA143780 zo31b03.s1 Stratagene
gb_est35:AI813806	+	37.00	128.04	229.14	665	! AI813806 wk79g12.x1 NCI_CGAP_Pa
gb_gss8:B96290	-	37.00	127.10	258.70	751	! B96290 T24D5TFB TAMU Arabidopsi
gb_est24:AI196751	-	36.00	135.40	89.16	157	! AI196751 ui66g07.x1 Sugano mous
gb_gss11:AQ319711	+	36.00	133.60	112.38	198	! AQ319711 RPCI11-105J20.TJ RPCI-
gb_est39:AV209680	+	36.00	131.85	140.69	248	! AV209680 AV209680 RIKEN full-le

gb_est31:AV045270   36.00   131.63   144.65   255 ! AV045270 AV045270 Mus muscul
gb_est33:AI768731   +   36.00   130.52   166.72   294 ! AI768731 wh25d12.x1 NCI_CGAP
gb_est16:AA594208   -   36.00   129.39   192.74   340 ! AA594208 nn29e02.s1 NCI_CGAP

seq_name: gb_est8:C03945

seq_documentation_block:
LOCUS   C03945   232 bp   mRNA   EST   30-JUL-1996
DEFINITION C03945 Human heart cDNA (YNakamura) Homo sapiens cDNA clone
            3NHC2454, mRNA sequence.
ACCESSION   C03945
VERSION   C03945.1 GI:1467196
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 232)
  AUTHORS   Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and
            Nakamura,Y.
  TITLE   Construction of a normalized directionally cloned cDNA library from
            adult heart and analysis of 3040 clones by partial sequencing
  JOURNAL   Genomics 35 (1), 231-235 (1996)
  MEDLINE   96299762
COMMENT   On Oct 24, 1995 this sequence version replaced gi:1040105.
            Contact: Yusuke Nakamura
            Institute of Medical Science
            University of Tokyo
            4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
            Tel: 81-3-5449-5372
            Fax: 81-3-5449-5433
            Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES   Location/Qualifiers
     source   1..232
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="3NHC2454"
                     /clone_lib="Human heart cDNA (YNakamura)"
                     /dev_stage="adult"
                     /note="Organ: heart; normalized directionally cloned cDNA
                     from adult heart"
BASE COUNT   55 a   77 c   68 g   32 t
ORIGIN

alignment_scores:
       Quality: 44.00   Length: 10
       Ratio: 4.889   Gaps: 0
       Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x C03945

Align seg 1/1 to: C03945   from: 1 to: 232

           1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
             |||||||||||||||||||| |||||||||
          40 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 69

seq_name: gb_est10:AA151891

seq_documentation_block:
LOCUS   AA151891   255 bp   mRNA   EST   10-DEC-1996
DEFINITION AA151891 zo01f06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:566435 5' similar to gb:M15497_cds1 HLA CLASS I
            HISTOCOMPATIBILITY ANTIGEN, A-24.(A-9) A*2401 (HUMAN):, mRNA
            sequence.
ACCESSION   AA151891
VERSION   AA151891.1 GI:1720754
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE
AUTHORS   Eutheria; Primates; Catarrhini; Hominidae; Homo.
             1 (bases 1 to 255)
             Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
             Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
             Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
             Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
             Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
             Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
             and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
COMMENT   On May 8, 1995 this sequence version replaced gi:800234.
             Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Trace considered overall poor quality
             Seq primer: -28M13 rev2 from Amersham
             High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source 1..255
             /organism="Homo sapiens"
             /db_xref="GDB:450888"
             /db_xref="taxon:9606"
             /clone="IMAGE:566435"
             /clone_lib="SOLR cells (kanamycin resistant)"
             /note="Organ: colon; Vector: pBluescript SK-; Site_1:
             EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
             Oligo dT. T-84 colonic epithelial cell line. Average
             insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
             sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
             CTCGAGTTTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT   57 a   70 c   75 g   44 t   9 others
ORIGIN

alignment_scores:
         Quality: 44.00   Length: 10
           Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x AA151891
Align seg 1/1 to: AA151891 from: 1 to: 255

  1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
    |||||||||||||||| ||||||||
 77 CGAGAGAACCTGCGGATCGGCNCTCCGCTAC 106

seq_name: gb_est11:AA263158

seq_documentation_block:
LOCUS   AA263158   283 bp   mRNA   EST   02-JUL-1998
DEFINITION PMY0534 KG1-a Lambda Zap Express cDNA library Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA263158
VERSION   AA263158.1 GI:1898964
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 283)
AUTHORS   Claudio,J.O., Liew,C.C., Dempsey,A.A., Cukerman,E., Stewart,A.K.,
            Na,E., Atkins,H.I., Iscove,N.N. and Hawley,R.G.
TITLE   Identification of sequence-tagged transcripts differentially

---

             expressed within the human hematopoietic hierarchy
JOURNAL   Genomics 50 (1), 44-52 (1998)
MEDLINE   9892493
COMMENT   On May 5, 1995 this sequence version replaced gi:797810.
             Contact: Hawley RG
             Oncology Research Laboratories
             The Toronto Hospital
             CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
             Tel: 416 3403834
             Fax: 416 3403453
             Email: r.hawley@utoronto.ca
             Similar to M58636 MHC class I HLA-Bw gene. Clone was randomly
             picked from KG1a primary library.
             Seq primer: 5' GAATTAACCCTCACTAAGGG 3'
             High quality sequence stop: 283.
FEATURES   Location/Qualifiers
     source 1..283
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="KG1-a Lambda Zap Express cDNA library"
             /cell_type="promyeloblast"
             /cell_line="KG1-a"
             /note="Vector: Lambda Zap Express (Stratagene); Site_1:
             EcoRI; Site_2: XhoI; Unidirectional cloning sites:
             EcoRI-XhoI. mRNA was purified from KG1-a cell line, cDNA
             was synthesized using an XhoI-OligodT linker primer. EcoRI
             adaptors were ligated, followed by digestion with XhoI for
             directional cloning into predigested Lambda Zap Express"
BASE COUNT   64 a   91 c   88 g   40 t
ORIGIN

alignment_scores:
         Quality: 44.00   Length: 10
           Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x AA263158
Align seg 1/1 to: AA263158 from: 1 to: 283

  1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
    |||||||||||||| ||||||||
120 CGAGAGAACCTGCGGATCGGCCGCTCCGCTAC 149

seq_name: gb_est6:D82221

seq_documentation_block:
LOCUS   D82221   375 bp   mRNA   EST   09-FEB-1996
DEFINITION HUMHBC4626 Human pancreatic islet Homo sapiens cDNA similar to
            HLA-B, mRNA sequence.
ACCESSION   D82221
VERSION   D82221.1 GI:1183739
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 375)
AUTHORS   Takeda,J.
TITLE   Human pancreatic islet ESTs
JOURNAL   Unpublished (1995)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:785255.
            Contact: Jun Takeda
            Institute for Molecular and Cellular Regulation, Gunma University
            3-39-15 Showa-machi, Maebashi Gunma 371, Japan
            Tel: 272-20-8856
            Fax: 272-20-8896
            Email: jtakeda@sb.gunma-u.ac.jp.
FEATURES   Location/Qualifiers
     source 1..375
             /organism="Homo sapiens"

```
                /db_xref="taxon:9606"
                /clone_lib="Human pancreatic islet"
                /note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
                I; mRNA was prepared from normal adult human islets. cDNA
                was directionally synthesized from the Xho I in the vector
                to the EcoRI site. cDNA was size fractionated to remove
                sequences <1000 bp in size."
BASE COUNT      75 a    124 c    118 g    55 t     3 others
ORIGIN

alignment_scores:
     Quality:     44.00              Length:        10
       Ratio:      4.889               Gaps:         0
Percent Similarity:  90.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x D82221
Align seg 1/1 to: D82221 from: 1 to: 375

     1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
       |||||||||||||||||||  ||||||||
   306 CGAGAGAACCTGCGGATCGCGCTCCGNTAC 335

seq_name: gb_est10:AA147151

seq_documentation_block:
LOCUS       AA147151      581 bp    mRNA            EST       05-DEC-1996
DEFINITION  zo32d06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:588587 5' similar to gb:M64740 HLA CLASS I HISTOCOMPATIBILITY
            ANTIGEN, A-24(A-9) A*2402 ALPHA (HUMAN);, mRNA sequence.
ACCESSION   AA147151
VERSION     AA147151.1  GI:1716526
KEYWORDS    EST.
SOURCE      human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE     Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393699.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 272.

FEATURES             Location/Qualifiers
     source          1..581
                     /organism="Homo sapiens"
                     /db_xref="GDB:4620889"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:588587"
                     /clone_lib="Stratagene colon (#937204)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. T-84 colonic epithelial cell line. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
```

```
                sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      134 a    162 c    185 g    85 t    15 others
ORIGIN

alignment_scores:
     Quality:     44.00              Length:        10
       Ratio:      4.889               Gaps:         0
Percent Similarity:  90.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x AA147151
Align seg 1/1 to: AA147151 from: 1 to: 581

     1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
       ||||||||||||||||||||  ||||||||
   152 CGAGAGAACCTGCGGATCNCGGCTCCGCTAC 181

seq_name: gb_est26:AI359260

seq_documentation_block:
LOCUS       AI359260      618 bp    mRNA            EST       15-FEB-1999
DEFINITION  qy27b07.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2013205 3'
            similar to gb:D32129 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
            AW-66(A-10) A*6601 ALPHA (HUMAN);, mRNA sequence.
ACCESSION   AI359260
VERSION     AI359260.1  GI:4110881
KEYWORDS    EST.
SOURCE      Homo sapiens
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 618)
AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
TITLE     Unpublished (1998)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT     Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 691    Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.

FEATURES             Location/Qualifiers
     source          1..618
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2013205"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCATATCTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
```

Align seg 1/1 to: AI696864 from: 1 to: 748

```
                1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
                  |||||||||||||||||   ||||||||||
              137 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 166
```

seq_name: gb_est13:AA332511

seq_documentation_block:
LOCUS   AA332511   360 bp   mRNA   EST   21-APR-1997
DEFINITION   EST36483 Embryo, 8 week I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA332511
VERSION   AA332511.1 GI:1984775
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 360)
AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
              Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
              White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
              Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
              Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
              Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
              Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
              Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
              Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
              Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
              Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
              Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
              He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
              Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
              Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
              Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
              Fraser,C.M. and Venter,J.C.
TITLE   Initial assessment of human gene diversity and expression patterns
              based upon 83 million nucleotides of cDNA sequence
JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE   12140200
COMMENT   On May 18, 1995 this sequence version replaced gi:811192.
              Other_ESTs: THC188992
              Contact: Kerlavage, AR
              Bioinformatics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850 USA
              Tel: 3018699056
              Fax: 3018699423
              Email: arkerlav@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST, please check the TIGR Human Gene
              Index (http://www.tigr.org/tdb/hgi/hgi.html)
              Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..360
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):134173"
                     /db_xref="taxon:9606"
                     /clone_lib="Embryo, 8 week I"
                     /dev_stage="Embryo, 8 wks"
                     /note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT   87 a   134 c   67 g   67 t   5 others
ORIGIN
```

alignment_scores:
   Quality:    40.00            Length:        9
   Ratio:       4.444            Gaps:          0
Percent Similarity: 100.000     Percent Identity: 88.889

alignment_block:
US-08-653-294-10 x AA332511

---

```
BASE COUNT       128 a      171 c      182 g      137 t
ORIGIN
```

alignment_scores:
   Quality: 44.00   Length: 10
   Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x AI359260/rev   ..

Align seg 1/1 to reverse of: AI359260 from: 1 to: 618

```
                1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
                  ||||||||||||||||||   |||||||||
              319 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 290
```

seq_name: gb_est31:AI696864

seq_documentation_block:
LOCUS   AI696864   748 bp   mRNA   EST   03-JUN-1999
DEFINITION   wc74h11.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2324421 3'
              similar to gb:M28205 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
              B-51(B-5) B*5101 ALPHA (HUMAN):, mRNA sequence.
ACCESSION   AI696864
VERSION   AI696864.1 GI:4984764
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 748)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Mar 16, 1998 this sequence version replaced gi:2961758.
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              Life Technologies catalog #: 11548-013
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www.bio.llnl.gov/bbrp/image/image.html

              Seq primer: -40UP from Gibco
              High quality sequence stop: 424.
FEATURES   Location/Qualifiers
     source   1..748
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2324421"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT   169 a   227 c   237 g   108 t   7 others
ORIGIN
```

alignment_scores:
   ,Quality: 44.00   Length: 10
   Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-10 x AI696864   ..

Align seg 1/1 to: AA332511 from: 1 to: 360

                    2 GluAsnLeuArgIleLeuArgTyr 10
                      ||||||||||||||||||||||||
                  333 GAAAATTGCGGATTCTGCTTAGGTTT 359

seq_name: gb_est38:AL036690

seq_documentation_block:
LOCUS   AL036690   171 bp   mRNA   EST   27-SEP-1999
DEFINITION   DKFZp564D2463_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
             DKFZp564D2463 5', mRNA sequence.
ACCESSION   AL036690
VERSION   AL036690.3 GI:5927859
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 171)
AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
             Wiemann,S.
TITLE   EST (Duesterhoeft, et al.)
JOURNAL   Unpublished (1999)
COMMENT   On Jul 7, 1999 this sequence version replaced gi:5866258.
             Contact: Duesterhoeft A
             MIPS
             Am Klopferspitz 18a D-82152 Martinsried, Germany
             This is the 5' sequence of the clone insert
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
             sequenced by Qiagen within the cDNA sequencing consortium of the
             German Genome Project.
             No s1 sequence available.
             This clone is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..171
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp564D2463"
                     /clone_lib="564 (synonym: hfbr2)"
                     /tissue_type="brain"
                     /dev_stage="fetal"
                     /lab_host="X1-2blue"
                     /note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT   36 a   53 c   60 g   22 t
ORIGIN

alignment_scores:
          Quality: 39.00   Length: 10
          Ratio: 4.333   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 80.000

alignment_block:
US-08-653-294-10 x AL036690   ..

Align seg 1/1 to: AL036690 from: 1 to: 171

                    1 ArgGluAsnLeuArgIleLeuArgTyr 10
                      ||||||::||||||| |||||||||||
                   62 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 91

seq_name: gb_gss9:AQ163964

seq_documentation_block:
LOCUS   AQ163964   351 bp   DNA   GSS   16-OCT-1998
DEFINITION   HS_2270_B1_A09_MF CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate=2270 Col=17 Row=B, genomic survey
             sequence.

ACCESSION   AQ163964
VERSION   AQ163964.1 GI:3552159
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 351)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Sequence Tagged Connector
             Plate: 2270 row: B column: 17
             Class: BAC ends
             High quality sequence stop: 351.
FEATURES   Location/Qualifiers
     source   1..351
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-2270 Col-17 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT   123 a   52 c   47 g   128 t   1 others
ORIGIN

alignment_scores:
          Quality: 39.00   Length: 9
          Ratio: 4.333   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-10 x AQ163964   ..

Align seg 1/1 to: AQ163964 from: 1 to: 351

                    2 GluAsnLeuArgIleLeuArgTyr 10
                      :::||||||||||||:::||||||||
                  292 GATAATTAAGAATTATACTTAGATAT 318

seq_name: gb_est23:AI124815

seq_documentation_block:
LOCUS   AI124815   414 bp   mRNA   EST   11-SEP-1998
DEFINITION   am56e06.x1 Johnston frontal cortex Homo sapiens cDNA clone
             IMAGE:1539586 3' similar to gb:M24038_cds1 HLA CLASS I
             HISTOCOMPATIBILITY ANTIGEN, BW-44(B-12) B*4402 (HUMAN);contains
             MER22.t3 TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI124815
VERSION   AI124815.1 GI:3593329
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 414)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
             Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
             Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
             Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE   WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   On Jan 17, 1998 this sequence version replaced gi:1899887.
             Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Seq primer: -40m13 fwd ET from Amersham.
FEATURES   Location/Qualifiers
     source   1..414
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1539586"
                     /clone_lib="Johnston frontal cortex"
                     /sex="male"
                     /tissue_type="pooled frontal lobe"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: brain; Vector: Bluescript SK-; Site_1;
                     EcoRI; Stanley Neuropathology Consortium
                     (www.stanleylab.org) brains S-58, S-65, S-67, S-78.
                     Random + oligo-dT primed into EcoRI site of ZAP II Vector.
                     Mass excised. Avg insert length 1.9kb. Custom library
                     provided by Dr. Nancy Johnston [(410) 614-3918,
                     nlj@welchlink.welch.jhu.edu]."
BASE COUNT   80 a   140 c   136 g   58 t
ORIGIN

alignment_scores:
     Quality: 39.00   Length: 10
     Ratio: 4.875   Gaps: 0
Percent Similarity: 80.000   Percent Identity: 80.000

alignment_block:
US-08-653-294-10 x AI124815

Align seq 1/1 to: AI124815 from: 1 to: 414

   1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
     |||||||||||| ||||||||
 284 CGAGAGAACCTGCGCACCGGCGTCCGCCTAC 313

seq_name: gb_gss15:AQ614213

seq_documentation_block:
LOCUS   AQ614213   498 bp   DNA   GSS   15-JUN-1999
DEFINITION HS_5123_B1_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=699 Col=11 Row=D, genomic survey sequence.
ACCESSION   AQ614213
VERSION   AQ614213.1 GI:5075489
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 498)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE   Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618

            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web server:
            http://www.htsc.washington.edu
            Plate: 699 row: D column: 11
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 498.
FEATURES   Location/Qualifiers
     source   1..498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=699 Col=11 Row=D"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT   129 a   121 c   98 g   150 t
ORIGIN

alignment_scores:
     Quality: 39.00   Length: 9
     Ratio: 4.333   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-10 x AQ614213

Align seq 1/1 to: AQ614213 from: 1 to: 498

   2 GluAsnLeuArgIleLeuLeuArgTyr 10
     :::|||||||||::::||||||||||||
 206 CAGAATCTTCGCTTGTTACTGCGATAT 232

seq_name: gb_est21:AA989542

seq_documentation_block:
LOCUS   AA989542   402 bp   mRNA   EST   02-JUN-1998
DEFINITION am64d02.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
            IMAGE:1576803 3' similar to gb:L05093 60S RIBOSOMAL PROTEIN L18A
            (HUMAN);, mRNA sequence.
ACCESSION   AA989542
VERSION   AA989542.1 GI:3174906
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 402)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE   WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2153091.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: -40m13 fwd ET from Amersham
            High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..402

```
                                        Ratio:    4.750       Gaps:              0
                   Percent Similarity:  80.000   Percent Identity:  80.000

                   alignment_block:
                   US-08-653-294-10 x D82189     ..

                   Align seg 1/1  to: D82189  from: 1  to: 415

                        1 ArgGluAsnLeuArgIleLeuLeuArgTyr 10
                          |||||||||||||  ||| ||||||||
                      279 CGAGAGAACCTGNGATCGCGCTCCGNTAC 308

                   seq_name: gb_est38:AW092686

                   seq_documentation_block:
                   LOCUS       AW092686        440 bp    mRNA                     EST     18-OCT-1999
                   DEFINITION  EST285866 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
                               clone cLET21F6, mRNA sequence.
                   ACCESSION   AW092686
                   VERSION     AW092686.1  GI:6058281
                   KEYWORDS    EST.
                   SOURCE      tomato.
                   ORGANISM    Lycopersicon esculentum
                               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                               euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
                               eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
                               Potatoe; Lycopersicon.
                   REFERENCE   1  (bases 1 to 440)
                   AUTHORS     D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
                               Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
                               Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and
                               Giovannoni,J.
                   TITLE       Generation of ESTs from tomato leaf tissue
                   JOURNAL     Unpublished (1999)
                   COMMENT     On Jun 22, 1998 this sequence version replaced gi:3246706.
                               Contact: David Frisch
                               Clemson University Genomics Institute
                               Clemson University
                               100 Jordan Hall, Clemson, SC 29634, USA
                               Tel: 864 656 4366
                               Fax: 864 656 4293
                               Email: dfrisch@CLEMSON.EDU
                               5 prime sequence.
                   FEATURES             Location/Qualifiers
                        source          1..440
                                        /organism="Lycopersicon esculentum"
                                        /cultivar="Rio Grande PtoR"
                                        /db_xref="taxon:4081"
                                        /clone="cLET21F6"
                                        /clone_lib="tomato mixed elicitor, BTI"
                                        /tissue_type="leaf"
                                        /dev_stage="4-6 week old plants"
                                        /lab_host="XL1-Blue MRF'"
                                        /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
                                        XhoI; cLET - Inoculated with a variety of disease response
                                        elicitors. Plants exposed to 2,6 dichloroisonicotinic
                                        acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
                                        okadaic acid, or systemin prior to tissue harvest. EcoRI
                                        site was destroyed during cloning."
                   BASE COUNT     112 a      102 c      96 g      130 t
                   ORIGIN

                   alignment_scores:
                        Quality:   38.00                      Length:         10
                          Ratio:    4.222                       Gaps:          0
                   Percent Similarity:  90.000   Percent Identity:  80.000

                   alignment_block:
                   US-08-653-294-10 x AW092686/rev  ..

                   Align seg 1/1  to reverse of: AW092686  from: 1  to: 440
```

```
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:1576803"
                   /clone_lib="Barstead spleen HPLRB2"
                   /sex="male"
                   /dev_stage="adult, 17 years"
                   /lab_host="DH10B"
                   /note="Organ: spleen; Vector: pT7T3D-Pac (Pharmacia) with
                   a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                   strand cDNA was primed with a Not I - oligo(dT) primer [5'
                   TGTTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT
                   3']; double-stranded cDNA was ligated to Eco RI adaptors
                   [AATTCGGATCCTTG], digested with Not I and cloned into the
                   Not I and Eco RI sites of the modified pT7T3 vector.
                   Library constructed by Bob Barstead. "
BASE COUNT      72 a      106 c     132 g      92 t
ORIGIN

alignment_scores:
     Quality:   38.00                      Length:         9
       Ratio:    4.222                       Gaps:          0
Percent Similarity: 100.000   Percent Identity:  88.889

alignment_block:
US-08-653-294-10 x AA989542/rev  ..

Align seg 1/1  to reverse of: AA989542  from: 1  to: 402

     2 GluAsnLeuArgIleLeuLeuArgTyr 10
       ||:::|||||||||||||||||
   353 GAAGAACTCCGGATCTTGCTGCGCTAT 327

seq_name: gb_est6:D82189

seq_documentation_block:
LOCUS       D82189          415 bp    mRNA                     EST     09-FEB-1996
DEFINITION  HUMHBC4524 Human pancreatic islet Homo saplens cDNA similar to
            HLA-B, mRNA sequence.
ACCESSION   D82189
VERSION     D82189.1  GI:1183662
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
AUTHORS     Takeda,J.
TITLE       Human pancreatic islet ESTs
JOURNAL     Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785206.
            Contact: Jun Takeda
            Institute for Molecular and Cellular Regulation, Gunma University
            3-39-15 Showa-machi, Maebashi Gunma 371, Japan
            Tel: 272-20-8856
            Fax: 272-20-8896
            Email: jtakeda@sb.gunma-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1..415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human pancreatic islet"
                     /note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
                     I; mRNA was prepared from normal adult human islets. cDNA
                     was directionally synthesized from the Xho I in the vector
                     to the EcoRI site. cDNA was size fractionated to remove
                     sequences <1000 bp in size."
BASE COUNT      79 a      134 c     133 g      61 t       8 others
ORIGIN

alignment_scores:
     Quality:   38.00                      Length:        10
```

us-08-653-294-10.rst

```
  1 ArgGluAsnLeuArgIleLeuLeuArgTyr  10
    |||||||||||   ||||||::::|||||
142 AGGGAGAATTAGTAATTCTTGTACGTTAT  113
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:29:37 ; Search time 122.56 Seconds
                 (without alignments)
                 1.933 Million cell updates/sec

Title: US-08-653-294-11
Perfect score: 49
Sequence: 1 YRLLIRLNER 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	49	100.0	10	1	W47268	Immunomodulatory p
2	44	89.8	10	1	W47266	Immunomodulatory p
3	44	89.8	10	1	W47272	Immunomodulatory p
4	44	89.8	20	1	R92909	HLA-B2702 CTL modu
5	44	89.8	20	1	R92911	HLA-B2702 CTL modu
6	44	89.8	20	1	R92907	HLA-B2702 CTL modu
7	44	89.8	20	1	R95428	HLA-B2702 84-75-84
8	44	89.8	20	1	W33778	Immunomodulating d
9	44	89.8	20	1	W33779	Immunomodulating d
10	44	89.8	20	1	W33792	Peptide B2702.84-7
11	39	79.6	10	1	W47270	Immunomodulatory p
12	39	79.6	20	1	R92910	HLA-B2702 CTL modu
13	39	79.6	20	1	R92908	HLA-B2702 CTL modu
14	39	79.6	20	1	R95430	HLA-B2702 84-75T/7
15	39	79.6	20	1	W33791	Peptide B2702.84-7
16	39	79.6	20	1	W33793	Peptide B2702.84-7
17	34	69.4	239	1	W74405	S. aureus gidB pro
18	34	69.4	239	1	W74406	S. aureus gidB pro
19	34	69.4	383	1	W41592	Rat FRAG1 protein.
20	34	69.4	485	1	R20796	EHV-4 gC. Nucleic
21	32	65.3	230	1	R63199	Temp. sensitive au
22	30	63.3	3079	1	R59926	GAP protein Ira2.
23	30	61.2	495	1	W89551	Mouse Smad6 protei
24	30	61.2	496	1	W96815	Smad7 protein used
25	30	61.2	721	1	W34454	Racillus subtilis
26	30	61.2	746	1	W34455	Racillus subtilis
27	30	61.2	851	1	R41333	113 kD ISGF-3alpha
28	30	61.2	851	1	R72077	Recognition factor
29	30	61.2	851	1	W03166	Human STAT2. New S
30	29	59.2	6	1	W47264	Immunomodulatory p
31	29	59.2	6	1	W33783	Peptide #4 used in
32	29	59.2	81	1	R62906	TCR alpha chain of
33	29	59.2	90	1	Y11975	Human 5' EST secre
34	29	59.2	92	1	R62904	TCR alpha-chain va
35	29	59.2	92	1	R62905	TCR alpha-chain va
36	29	59.2	92	1	R62907	V-alpha-15 gene al
37	29	59.2	112	1	W98457	H. pylori GHPO 669
38	29	59.2	126	1	R26075	Mutant E2 binding
39	29	59.2	156	1	W56254	Interleukin-13 bin
40	29	59.2	183	1	W05730	Gene 71 product co
41	29	59.2	389	1	R15428	3-acylation enzyme
42	29	59.2	391	1	R97659	Single chain T cel
43	29	59.2	464	1	W82652	Ehrlichia sp. exte
44	29	59.2	706	1	W31199	Bacillus popilliae
45	28	57.1	21	1	W11013	Antigen peptide de

ALIGNMENTS

RESULT   1
W47268
ID   W47268 standard; peptide; 10 AA.
AC   W47268;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence 10 AA;

  Query Match   100.0%; Score 49; DB 1; Length 10;
  Best Local Similarity 100.0%; Pred. No. 0.00023;
  Matches   10; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 YRLLIRLNER 10
            ||||||||||
Db   1 YRLLIRLNER 10

RESULT   2
W47266
ID   W47266 standard; peptide; 10 AA.
AC   W47266;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   Misc_difference 1..10

```
FT              /note= "at least one of the amino acids is the
FT                      D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI: 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence   10 AA;

Query Match         89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity   90.0%;  Pred. No. 0.0025;
Matches    9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT    3
W47272
ID   W47272 standard; peptide; 10 AA.
AC   W47272;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key            Location/Qualifiers
FT   Misc_difference 1..10
FT              /note= "at least one of the amino acids is the
FT                      D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI: 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence   10 AA;

Query Match         89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity   90.0%;  Pred. No. 0.0025;
Matches    9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        |||||||:|:||
Db    1 YRLLIRLDER 10

RESULT    4
R92909
ID   R92909 standard; peptide; 20 AA.
AC   R92909;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC:
KW   immunosuppressant; graft versus host disorder; transplantation; therapy:
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO952679-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C,  Krensky AM,  Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major.histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match         89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity   90.0%;  Pred. No. 0.0054;
Matches    9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        ||| |||||||
Db    1 YRLAIRLNER 10

RESULT    5
R92911
ID   R92911 standard; peptide; 20 AA.
AC   R92911;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/84-75).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy:
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO952679-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C,  Krensky AM,  Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is a dimer of residues 75-84 of the alpha-1 domain of the class I MHC.
```

```
CC  HLA-B2702. These sequences can be used to extend the period of
CC  acceptance by a recipient of a transplant from an MHC unmatched donor.
CC  The peptides are administered to a patient in conjunction with a
CC  subtherapeutic amount of an immunosuppressant. This is administered to
CC  the patient for a limited period of time (compared to the lifetime
CC  administration for current treatments). The peptides particularly
CC  modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC  of the patient.
SQ  Sequence 20 AA.

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity 90.0%;  Pred. No. 0.0054;
Matches   9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT 6
R92907
ID  R92907 standard; peptide; 20 AA.
AC  R92907;
DT  16-MAY-1996 (first entry)
DE  HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW  Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW  immunosuppressant; graft versus host disorder; transplantation; therapy;
KW  class I MHC; HLA-B2702.
OS  Synthetic.
PN  WO9526979-A1.
PD  12-OCT-1995.
PF  05-APR-1995; U04349.
PR  05-APR-1994; US-222851.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Clayberger C, Krensky AM, Parham P;
DR  WPI: 95-358582/46.
PT  Extension of acceptance period of transplants from MHC unmatched
PT  donor hosts - using Class I B75-84 MHC antigen of the recipient
PT  host
PS  Example 15; Page 36; 80pp; English.
CC  R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC  class I major histocompatibility complex (MHC) antigens. This sequence
CC  is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC  I MHC HLA-B2702. These sequences can be used to extend the period of
CC  acceptance by a recipient of a transplant from an MHC unmatched donor.
CC  The peptides are administered to a patient in conjunction with a
CC  subtherapeutic amount of an immunosuppressant. This is administered to
CC  the patient for a limited period of time (compared to the lifetime
CC  administration for current treatments). The peptides particularly
CC  modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC  of the patient.
SQ  Sequence 20 AA.

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity 90.0%;  Pred. No. 0.0054;
Matches   9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT 7
R95428
ID  R95428 standard; peptide; 20 AA.
AC  R95428;
DT  12-NOV-1996 (first entry)
DE  HLA-B2702 84-75-84 palindrome.
KW  HLA; B2702; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW  T-cell lysate; membrane protein; mammal; heat shock protein: Hsc70; APC;
KW  B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;

KW  cytolysis; antigen presenting cell.
OS  Synthetic.
PN  WO9513288-A1.
PD  18-MAY-1995.
PF  10-NOV-1994; U12985.
PR  10-NOV-1993; US-150493.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Clayberger C, Krensky AM;
DR  WPI: 95-194027/25.
PT  Compsns. comprising lymphoid surface membrane proteins - which may
PT  inhibit cytolytic activity and differentiation of CTLs.
PS  Example; Page 12; 29pp; English.
CC  R95413, and R95415-R95431 represent palindromes and fragments of
CC  human-leucocyte-associated antigens. This sequence represents the
CC  HLA-B2702 84-75-84 palindrome. These sequences can be used to isolate
CC  the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC  protein associated with T-cell activation in mammalian T-cells, and is
CC  also immunologically cross reactive with the heat shock protein Hsc70.
CC  p74 is found in a limited number of cell types, but is particularly
CC  expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC  cell with an amphoteric detergent, and then passed through an affinity
CC  column containing a covalently bound HLA-B2702 palindromic peptide.
CC  Compositions comprising the extracellular fragment of p74 combined with
CC  HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC  cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC  compounds can be screened for their effect on the cytolytic activity of
CC  T-cells, by combining them with the extracellular portion of p74 and
CC  determining the amount of binding between the candidate compound and p74.
CC  Modulation of CTL activity can be inhibited in a cellular composition
CC  containing T-cells and antigen presenting cells (APCs), by adding to the
CC  mix the extracellular portion of p74, in an amount sufficient to compete
CC  with p74 for the binding of the p74 ligand.
SQ  Sequence 20 AA.

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity 90.0%;  Pred. No. 0.0054;
Matches   9;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT 8
W33778
ID  W33778 standard; peptide; 20 AA.
AC  W33778;
DT  19-JUN-1998 (first entry)
DE  Immunomodulating dimer peptide #1.
KW  Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW  transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW  rejection.
OS  Homo sapiens.
OS  Synthetic.
PN  WO9745351-A1.
PD  27-NOV-1997.
PF  22-MAY-1997; U08689.
PR  24-MAY-1996; US-653294.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Beulow R, Clayberger C, Krensky AM;
DR  WPI: 98-086530/08.
PT  New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT  alpha-1 domain, used for preventing rejection of transplants or
PT  treating autoimmune diseases
PS  Claim 16; Page 35; 41pp; English.
CC  This sequence represents a specifically claimed immunomodulating
CC  dimer peptide of the invention. A peptide-type compound or variant is
CC  claimed which has immunomodulating activity, including the N-terminal
CC  acylated and/or C-terminal amidated or esterified forms of up to 60
CC  amino acids, where the peptide-type compound comprises the formula; A-B,
CC  where A, B -(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or
CC  V: aa77 - D, S or N; aa79 - R or G; aa80 = I or N; aa81, aa84- a
```

```
CC  hydrophobic or small amino acid; aa82 - R or L; aa83 - G or R; and aa
CC  represents amino acid. The sequence in the brackets may optionally be
CC  absent or truncated at any peptide type bond within the brackets. The
CC  compounds comprise amino acid sequences related to a Class I HLA-B
CC  alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC  T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC  vitro. They can also be used in combination with antigenic peptides or
CC  proteins of interest to activate CTLs. They can also inhibit the
CC  proliferation of T cells in response to anti-CD3. The peptide can be
CC  used for preventing rejection of transplants or for treating autoimmune
CC  diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC  The products can also be used for detection and diagnosis.
SQ  Sequence 20 AA;

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity  90.0%;  Pred. No. 0.0054;
Matches  9;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT    9
W33779
ID  W33779 standard; peptide; 20 AA.
AC  W33779;
DT  19-JUN-1998 (first entry)
DE  Immunomodulating dimer peptide #2.
KW  Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW  transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW  rejection.
OS  Synthetic.
OS  Homo sapiens.
PN  WO9744351-A1.
PD  27-NOV-1997.
PF  22-MAY-1997; U08689.
PR  24-MAY-1996; US-653294.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Beulow R, Clayberger C, Krensky AM;
DR  WPI; 98-086530/08.
PT  New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT  alpha-1 domain, used for preventing rejection of transplants or
PT  treating autoimmune diseases
PS  Claim 16; Page 35; 41pp; English.
CC  This sequence represents a specifically claimed immunomodulating
CC  dimer peptide of the invention. A peptide-type compound or variant is
CC  claimed which has immunomodulating activity, including the N-terminal
CC  acylated aad/or C-terminal amidated or esterified forms of up to 60
CC  amino acids, where the peptide-type compound comprises the formula; A-B,
CC  where A, B -(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or
CC  V; aa77 - D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a
CC  hydrophobic or small amino acid; aa82 - R or L; aa83 - G or R; and aa
CC  represents amino acid. The sequence in the brackets may optionally be
CC  absent or truncated at any peptide type bond within the brackets. The
CC  compounds comprise amino acid sequences related to a Class I HLA-B
CC  alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC  T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC  vitro. They can also be used in combination with antigenic peptides or
CC  proteins of interest to activate CTLs. They can also inhibit the
CC  proliferation of T cells in response to anti-CD3. The peptide can be
CC  used for preventing rejection of transplants or for treating autoimmune
CC  diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC  The products can also be used for detection and diagnosis.
SQ  Sequence 20 AA;

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity  90.0%;  Pred. No. 0.0054;
Matches  9;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT    10
W33792
ID  W33792 standard; peptide; 20 AA.
AC  W33792;
DT  19-JUN-1998 (first entry)
DE  Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW  Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW  transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW  rejection.
OS  Synthetic.
OS  Homo sapiens.
PN  WO9744351-A1.
PD  27-NOV-1997.
PF  22-MAY-1997; U08689.
PR  24-MAY-1996; US-653294.
PA  (STRD ) UNIV LELAND STANFORD JUNIOR.
PI  Beulow R, Clayberger C, Krensky AM;
DR  WPI; 98-086530/08.
PT  New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT  alpha-1 domain, used for preventing rejection of transplants or
PT  treating autoimmune diseases
PS  Example 1; Page 19; 41pp; English.
CC  Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC  activity. A peptide-type compound or variant is claimed which has
CC  immunomodulating activity, including the N-terminal acylated and/or
CC  C-terminal amidated or esterified forms of up to 60 amino acids, where
CC  the peptide-type compound comprises the formula; A-B, where A, B -
CC  (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or V, aa77 -
CC  D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a hydrophobic or
CC  small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC  acid. The sequence in the brackets may optionally be absent or truncated
CC  at any peptide type bond within the brackets. The compounds comprise
CC  amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC  79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC  undesirably attacking cells in a host or in vitro. They can also be
CC  used in combination with antigenic peptides or proteins of interest to
CC  activate CTLs. They can also inhibit the proliferation of T cells in
CC  response to anti-CD3. The peptide can be used for preventing rejection
CC  of transplants or for treating autoimmune diseases; e.g diabetes,
CC  rheumatoid arthritis and lupus erythematosis. The products can also be
CC  used for detection and diagnosis.
SQ  Sequence 20 AA;

Query Match     89.8%;  Score 44;  DB 1;  Length 20;
Best Local Similarity  90.0%;  Pred. No. 0.0054;
Matches  9;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy    1 YRLLIRLNER 10
        ||| ||||||
Db    1 YRLAIRLNER 10

RESULT    11
W47270
ID  W47270 standard; peptide; 10 AA.
AC  W47270;
DT  22-MAY-1998 (first entry)
DE  Immunomodulatory peptide.
KW  Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW  transplant rejection; treatment; autoimmune disease.
OS  Homo sapiens.
OS  Synthetic.
FH  Key           Location/Qualifiers
FT  Misc_difference 1..10
FT               /note= "at least one of the amino acids is the
FT                       D-isomer
PN  WO9744052-A1.
PD  27-NOV-1997.
```

PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence   10 AA;

Query Match   79.6%;   Score 39;   DB 1;   Length 10;
Best Local Similarity   80.0%;   Pred. No. 0.027;
Matches   8;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLNER 10
       ||| |||||:||
Db   1 YRLAIRLDER 10

RESULT 12
R92910
ID   R92910 standard; peptide; 20 AA.
AC   R92910;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match   79.6%;   Score 39;   DB 1;   Length 20;
Best Local Similarity   80.0%;   Pred. No. 0.059;
Matches   8;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLNER 10
       ||| |||||
Db   1 YRLATRLNER 10

RESULT 13
R92908
ID   R92908 standard; peptide; 20 AA.
AC   R92908;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match   79.6%;   Score 39;   DB 1;   Length 20;
Best Local Similarity   80.0%;   Pred. No. 0.059;   English.
Matches   8;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLNER 10
       ||| |||||
Db   1 YRLATRLNER 10

RESULT 14
R95430
ID   R95430 standard; peptide; 20 AA.
AC   R95430;
DT   12-NOV-1996 (first entry)
DE   HLA-B2702 84-75T/75-84T palindrome.
KW   HLA: p74: alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-75T/75-84T palindrome. These sequences can be used to
CC   isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface
CC   membrane protein associated with T-cell activation in mammalian T-cells,
CC   and is also immunologically cross reactive with the heat shock protein
CC   Hsc70. p74 is found in a limited number of cell types, but is
CC   particularly expressed on B and T cells. p74 can be isolated by lysis of

CC   a suitable cell with an amphoteric detergent, and then passed through an
CC   affinity column containing a covalently bound HLA-B2702 palindromic
CC   peptide. Compositions comprising the extracellular fragment of p74
CC   combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and
CC   inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.
CC   Candidate compounds can be screened for their effect on the cytolytic
CC   activity of T-cells, by combining them with the extracellular portion of
CC   p74 and determining the amount of binding between the candidate compound
CC   and p74. Modulation of CTL activity can be inhibited in a cellular
CC   composition containing T-cells and antigen presenting cells (APCs), by
CC   adding to the mix the extracellular portion of p74, in an amount
CC   sufficient to compete with p74 for the binding of the p74 ligand.
SQ   Sequence   20 AA;

Query Match   79.6%; Score 39; DB 1; Length 20;
Best Local Similarity   88.9%; Pred. No. 0.059;
Matches   8; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
              ||| |||||
Db   1 YRLAIRLNE 9

RESULT 15
W33791
ID   W33791 standard; peptide; 20 AA.
AC   W33791;
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75T/75-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B -
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or V; aa77 -
CC   D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a hydrophobic or
CC   small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   20 AA;

Query Match   79.6%; Score 39; DB 1; Length 20;
Best Local Similarity   80.0%; Pred. No. 0.059;
Matches   8; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNER 10
              ||| |||||
Db   1 YRLATRLNER 10

Search completed: February 8, 2000, 01:29:38
Job time: 1750 sec

```
              GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 7, 2000, 11:54:22 ; Search time 117.7 Seconds
                                         (without alignments)
                                         4.008 Million cell updates/sec

Title:          US-08-653-294-11
Perfect score:  49
Sequence:       1 YRLLIRLNER 10

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters:    142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :  PIR_62:*
            1:  pir1:*
            2:  pir2:*
            3:  pir3:*
            4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	73.5	395	2	T01392	leucine-rich repea
2	34	69.4	254	2	JC6168	fibroblast growth
3	34	69.4	485	1	B45343	glycoprotein gp13
4	33	67.3	201	2	S64994	probable membrane
5	33	67.3	348	2	I37271	cylicin II - human
6	33	67.3	468	1	B46114	glycoprotein gp13
7	33	67.3	506	2	S37583	RING finger protei
8	33	67.3	513	1	TVHURF	ret finger protein
9	33	67.3	560	2	I50372	ORF2 - chicken
10	33	67.3	801	4	TVHURE	transforming prote
11	33	67.3	1711	1	A47392	chromodomain-helic
12	32	65.3	151	2	C71113	probable frxA prot
13	32	65.3	183	2	B70744	hypothetical prote
14	32	65.3	221	2	S56263	hypothetical prote
15	32	65.3	239	2	S57158	dolichyl-phosphate
16	32	65.3	267	2	A32122	peptide transport
17	32	65.3	330	2	F64877	peptide transport
18	32	65.3	330	2	S39588	glycoprotein gp13
19	32	65.3	468	1	VGBEEH	probable penicilli
20	32	65.3	678	2	B70913	bud emergence prot
21	32	65.3	1128	2	A49960	hypothetical prote
22	31	63.3	214	2	T05004	hypothetical prote
23	31	63.3	297	2	T12615	ribosomal protein
24	31	63.3	386	2	T12048	ribosomal protein
25	31	63.3	397	2	T00914	leucine-rich repea
26	31	63.3	400	2	S07733	NADH dehydrogenase
27	31	63.3	425	2	T15959	hypothetical prote
28	31	63.3	825	2	S54465	YTA12 protein prec
29	31	63.3	880	1	B33926	DNA-directed RNA p
30	31	63.3	1131	2	T14517	hypothetical prote
31	31	63.3	2108	2	S28417	CDC39 protein - ye
32	31	63.3	3079	1	RGBYI2	probable GTPase-ac
33	30	61.2	144	2	A21047	ribosomal mobile e
34	30	61.2	173	2	S27599	hypothetical prote
35	30	61.2	176	2	F70007	hypothetical prote
36	30	61.2	182	1	TVFFR3	transforming prote
37	30	61.2	186	1	G70418	probable thiamine
38	30	61.2	206	2	E64327	H+-transporting AT
39	30	61.2	232	2	D75062	probable flagella-
40	30	61.2	242	2	B69664	N-acetylglucosamin
41	30	61.2	296	2	A33823	ribosomal protein
42	30	61.2	296	1	B33823	ribosomal protein
43	30	61.2	297	1	R5RTL5	ribosomal protein
44	30	61.2	297	2	S55912	ribosomal protein
45	30	61.2	297	2	JC1308	ribosomal protein

ALIGNMENTS

```
RESULT    1
T01392
leucine-rich repeat protein T4I9.11 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Mar-1999
C;Accession: T01392
R;Parnell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.; Habermann, K.; Schutz, K.;
submitted to the EMBL Data Library, May 1998
A;Description: Genomic sequence of BAC T4I9 from Arabidopsis thaliana, chromosome IV,
A;Reference number: 214314
A;Accession: T01392
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-395 <PAR>
A;Cross-references: EMBL:AF069442; NID:g3242970; PID:g3924604
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Note: T4I9.11

Query Match       73.5%;    Best Local Similarity   66.7%;    Score 36; DB 2; Length 395;
                                                              Pred. No. 7.3;
Matches   6;    Conservative   3;    Mismatches   0;    Indels  0;    Gaps  0;

Qy        1 YRLLIRLNE 9
            ||:||:|||
Db      231 YRVLLRLNQ 239


RESULT    2
JC6168
fibroblast growth factor receptor activating protein 1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
C;Accession: JC6168
R;Lorenzi, M.V.; Horii, Y.; Yamanaka, R.; Sakaguchi, K.; Miki, T.
Proc. Natl. Acad. Sci. U.S.A. 93, 8956-8961, 1996
A;Title: FRAG1, a gene that potently activates fibroblast growth factor receptor by C
A;Reference number: JC6168; MUID:96392347
A;Contents: osteosarcoma cell
A;Accession: JC6168
A;Molecule type: mRNA
A;Residues: 1-254 <LOR>
A;Cross-references: GB:U57715; NID:g1518608; PID:g1518609
A;Note: the authors translated the codon GTG for residue 56 as Cys
C;Comment: This protein plays an important role in cellular functions, and in fibrobl
C;Genetics:
A;Gene: frog1
C;Keywords: growth factor receptor; osteosarcoma

Query Match       69.4%;    Score 34; DB 2; Length 254;
```

Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 YRLLIRLN 8
       |||| |||
Db 109 YRLLCRLN 116

RESULT 3
B45343
glycoprotein gp13 precursor - equine herpesvirus 4
N:Alternate names: glycoprotein gC
C:Species: equine herpesvirus 4
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: B45343
R:Nicolson, L.; Onions, D.E.
  Virology 179, 378-387, 1990
A:Title: The nucleotide sequence of the equine herpesvirus 4 gC gene homologue.
A:Reference number: A45343; MUID:91021040
A:Accession: B45343
A:Molecule type: DNA
A:Residues: 1-485 <NIC>
A:Cross-references: GB:M58031; NID:g330894; PIDN:AAA46083.1; PID:g330896
C:Genetics:
C:Gene: 16
C:Superfamily: herpesvirus glycoprotein F
C:Keywords: glycoprotein; transmembrane protein <SIG>
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-485/Product: glycoprotein gp13 #status predicted <GGP>
F:60,61,66,67,72,108,116,147,220,225,286/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 69.4%; Score 34; DB 1; Length 485;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
       ||| ||| |
Db 120 YRLEIHLNQR 129

RESULT 4
S64994
probable membrane protein YLR145w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L3301
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 26-Aug-1999
C:Accession: S64994
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.
  submitted to the Protein Sequence Database, May 1996
A:Reference number: S64987
A:Accession: S64994
A:Molecule type: DNA
A:Experimental source: strain S288C
A:Residues: 1-201 <RIE>
A:Cross-references: EMBL:Z73317; NID:g1360569; PID:e245582; PIDN:g1360570; MIPS:YLR145w
C:Genetics:
C:Map position: 12R
C:Superfamily: Saccharomyces probable membrane protein YLR145w
C:Keywords: transmembrane protein
F:92-108/Domain: transmembrane protein #status predicted <TMM>

Query Match 67.3%; Score 33; DB 2; Length 201;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
       |||: ||: |||
Db 15 YRLILLNHR 24

RESULT 5

Best Local Similarity 67.3%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

I37271
cylicin II - human
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 25-Oct-1996
C:Accession: I37271; S52774
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.
  Exp. Cell Res. 218, 174-182, 1995
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads:
A:Reference number: I37271; MUID:95255491
A:Accession: I37271
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-348 <HES>
A:Cross-references: EMBL:Z46788; NID:g758586; PID:g758587

Query Match 67.3%; Score 33; DB 2; Length 348;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
       || |:||:||
Db 78 YRSLMRISER 87

RESULT 6
B46114
glycoprotein gp13 precursor - equine herpesvirus 1 (strain Kentucky A)
N:Alternate names: glycoprotein C
C:Species: equine herpesvirus 1
C:Note: host Equus caballus (domestic horse)
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Jul-1999
C:Accession: B46114
R:Matsumura, T.; Smith, R.H.; O'Callaghan, D.J.
  Virology 193, 910-923, 1993
A:Title: DNA sequence and transcriptional analyses of the region of the equine herpes
A:Reference number: A46114; MUID:93212524
A:Accession: B46114
A:Molecule type: DNA
A:Residues: 1-468 <MAT>
A:Cross-references: GB:S57839; NID:g298846; PIDN:AAB25944.1; PID:g298848
C:Superfamily: herpesvirus glycoprotein F
C:Keywords: glycoprotein; transmembrane protein
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-468/Product: glycoprotein gp13 #status predicted <GPT>
F:432-451/Domain: transmembrane #status predicted <TMN>
F:46,57,62,92,100,131,203,208,269/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 67.3%; Score 33; DB 1; Length 468;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
       || |:||:|
Db 104 YRLKIYLNQR 113

RESULT 7
S37583
RING finger protein rfp - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C:Accession: S37583
R:Takahashi, M.
  submitted to the EMBL Data Library, October 1993
A:Reference number: S37583
A:Accession: S37583
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-506 <TAK>
A:Cross-references: EMBL:X75343; NID:g406747; PIDN:CAA53092.1; PID:g406748
C:Superfamily: rfp transforming protein; RING finger homology

C:Keywords: zinc
F:5-55/Domain: RING finger homology <RNG>

Query Match   67.3%;   Score 33; DB 2; Length 506;
Best Local Similarity   77.8%;   Pred. No. 39;
  Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
          ||| ||| |||
Db   191 YRLLARLEE 199

RESULT   8
TVHURF
ret finger protein - human
N;Alternate names: transforming protein rfp
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 18-Jun-1999
C;Accession: A28101
R;Takahashi, M.; Inaguma, Y.; Hiai, H.; Hirose, F.
Mol. Cell. Biol. 8, 1853-1856, 1988
A;Title: Developmentally regulated expression of a human "finger"-containing gene encode
A;Reference number: A28101; MUID:88246464
A;Accession: A28101
A;Molecule type: mRNA
A;Residues: 1-513 <TAK>
A;Cross-references: DDBJ:J03407; NID:g337371; PIDN:AAA36564.1; PID:g337372
C;Genetics:
A;Gene: GDB:RFP
A;Cross-references: GDB:511359; GDB:1391662
A;Map position: 6p22-6p21.3
C;Superfamily: rfp transforming protein; RING finger homology
C;Keywords: DNA binding; transforming protein; zinc
F:1-315/Product: transforming protein rfp (fragment) #status predicted <RET>
F:12-52/Domain: RING finger homology <RNG>
F:16-127/Domain: metal and nucleic acid binding #status predicted <TMN>

Query Match   67.3%;   Score 33; DB 1; Length 513;
Best Local Similarity   77.8%;   Pred. No. 40;
  Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
          ||| ||| ||
Db   198 YRLLARLEE 206

RESULT   9
I50372
ORF2 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C;Accession: I50372
R;Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.
Development 119, 433-446, 1993
A;Title: Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomai
A;Reference number: I50372; MUID:94116444
A;Accession: I50372
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-560 <FUN>
A;Cross-references: GB:D14316; NID:g391639; PIDN:BAA03262.1; PID:g391640
C;Superfamily: CHD-1 protein; chromobox homology

Query Match   67.3%;   Score 33; DB 2; Length 560;
Best Local Similarity   77.8%;   Pred. No. 44;
  Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   2 RLLIRLNER 10
          :||||| ||
Db   43 KLLIRLRER 51

RESULT   10
TVHURE
transforming protein RFP/protein-tyrosine kinase RET mutant fusion protein - human
N;Alternate names: ret oncogene protein
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) ret
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 10-Sep-1997 #text_change 13-Aug-1999
C;Accession: A27203
R;Takahashi, M.; Cooper, G.M.
Mol. Cell. Biol. 7, 1378-1385, 1987
A;Title: ret transforming gene encodes a fusion protein homologous to tyrosine kinase
A;Reference number: A27203; MUID:87257826
A;Accession: A27203
A;Molecule type: mRNA
A;Residues: QAGA',1-801 <TAK>
A;Cross-references: GB:M16029; NID:g340025
A;Note: codons preceding the probable start codon were translated
C;Comment: The ret oncogene is the chimeric product of a translocation mutation betwe
C;Genetics:
A;Gene: RFP/RET
C;Keywords: ATP; fusion protein; oncogene; phosphotransferase; transforming protein;
F:1-315/Region: transforming protein rfp
F:316-792/Region: protein-tyrosine kinase ret
F:459-467/Region: protein kinase ATP-binding motif
F:487/Active site: Lys #status predicted

Query Match   67.3%;   Score 33; DB 4; Length 801;
Best Local Similarity   77.8%;   Pred. No. 63;
  Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
          ||| ||| |||
Db   198 YRLLARLEE 206

RESULT   11
A47392
chromodomain-helicase-DNA-binding protein, CHD-1 - mouse
N;Alternate names: KYBP protein
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A47392; S21568
R;Delmas, V.; Stokes, D.G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 2414-2418, 1993
A;Title: A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI
A;Reference number: A47392; MUID:93211972
A;Accession: A47392
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1711 <DEL>
A;Experimental source: S194 plasmacytoma cells
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:128272, NCBIP:128273)
R;Delmas, V.; Perry, R.P.
submitted to the EMBL Data Library, May 1992
A;Description: KYBP, a mammalian protein that contains the SNF2/SWI2 helicase domain
A;Reference number: S21568
A;Accession: S21568
A;Molecule type: mRNA
A;Residues: 772-1711 <DE2>
A;Cross-references: EMBL:X66028
C;Superfamily: CHD-1 protein; chromobox homology
C;Keywords: DNA binding
F:293-336/Domain: chromobox homology <CB1>
F:387-427/Domain: chromobox homology <CB2>

Query Match   67.3%;   Score 33; DB 1; Length 1711;
Best Local Similarity   77.8%;   Pred. No. 1.4e+02;
  Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

C;Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 05-Dec-1997
C;Accession: S56263
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccaromyces
A;Reference number: S56186
A;Accession: S56263
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-221 <MUR>
A;Cross-references: EMBL:D50617; NID:g836685; PID:d1009888; PID:g836763; MIPS:YFR008w
C;Genetics:
A;Map position: 6R

Query Match   65.3%;   Score 32;   DB 2;   Length 221;
Best Local Similarity   75.0%;   Pred. No. 27;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLN 8
          ||||:||:
Db   211 YRLLLRLH 218

RESULT   15
S57158
hypothetical protein YJR135c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein J2122
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 06-Feb-1998
A;Accession: S57158
R;Rose, M.; Koetter, P.; Entian, K.D.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56848
A;Accession: S57158
A;Molecule type: DNA
A;Residues: 1-239 <ROS>
A;Cross-references: EMBL:Z49635; NID:g1015871; PID:g1015872; MIPS:YJR135c
C;Genetics:
A;Gene: SGD:MCM22
A;Cross-references: SGD:S0003896; MIPS:YJR135c
A;Map position: 10R

Query Match   65.3%;   Score 32;   DB 2;   Length 239;
Best Local Similarity   75.0%;   Pred. No. 29;
Matches   6;   Conservative   1;   Mismatches   1;   Indels   1;   Gaps   0;

Qy   1 YRLLIRLN 8
          ||||:|| |
Db   205 YRLLLRAN 212

Search completed: February 7, 2000, 11:54:23
Job time: 24333 sec

Qy   2 RLLIRLNER 10
          :|||| ||
Db   793 KLLIRLER 801

RESULT   12
C71113
probable frxA protein - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
A;Accession: C71113
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: C71113
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-151 <KAW>
A;Cross-references: GB:AP000003; NID:g3236130; PID:d1030708; PID:g3257082
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0674

Query Match   65.3%;   Score 32;   DB 2;   Length 151;
Best Local Similarity   77.8%;   Pred. No. 18;
Matches   7;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   2 RLLIRLNER 10
          |||| :||
Db   96 RLLIELDER 104

RESULT   13
B70744
hypothetical protein Rv0487 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
A;Accession: B70744
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70744
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-183 <COL>
A;Cross-references: GB:Z77162; GB:AL123456; NID:g3261606; PID:e255029; PID:g1449285
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0487

Query Match   65.3%;   Score 32;   DB 2;   Length 183;
Best Local Similarity   60.0%;   Pred. No. 22;
Matches   6;   Conservative   1;   Mismatches   1;   Indels   3;   Gaps   0;

Qy   1 YRLLIRLNER 10
          || |:|| ||
Db   77 YRFLLRRNRR 86

RESULT   14
S55263
hypothetical protein YFR008w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: February 8, 2000, 00:59:51 ; Search time 63.71 Seconds
(without alignments)
4.688 Million cell updates/sec

Title: US-08-653-294-11
Perfect score: 49
Sequence: 1 YRLLIRLNER 10

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	34	69.4	485	1	VGLC_HSVE4	P22596 equine herp
2	33	67.3	299	1	RL5_BOMMO	O76190 bombyx mori
3	33	67.3	348	1	CYI2_HUMAN	Q14093 homo sapien
4	33	67.3	513	1	RFP_HUMAN	P14373 homo sapien
5	33	67.3	522	1	RFP_MOUSE	Q62158 mus musculu
6	33	67.3	1709	1	CHD1_HUMAN	O14646 homo sapien
7	33	67.3	1711	1	CHD1_MOUSE	P40201 mus musculu
8	32	65.3	183	1	Y487_MYCTU	Q11153 mycobacteri
9	32	65.3	221	1	YFH8_YEAST	P43592 saccharomyc
10	32	65.3	239	1	YJ9D_YEAST	P47167 saccharomyc
11	32	65.3	267	1	DPM1_YEAST	P14020 saccharomyc
12	32	65.3	327	1	RL5_ANOGA	O44248 anopheles g
13	32	65.3	330	1	SAPD_ECOLI	P36635 escherichia
14	32	65.3	330	1	SAPD_SALTY	P12889 equine herp
15	32	65.3	468	1	VGLC_HSVEB	P32873 saccharomyc
16	32	65.3	1128	1	BEM3_YEAST	P52822 schizosacch
17	31	63.3	294	1	RL5A_SCHPO	O74306 schizosacch
18	31	63.3	294	1	RL5B_SCHPO	O65353 helianthus
19	31	63.3	297	1	RL5_HELAN	P49165 urechis cau
20	31	63.3	386	1	RL4_URECA	P15689 paramecium
21	31	63.3	400	1	NUCM_PARTE	P40341 mus musculu
22	31	63.3	825	1	RCA1_YEAST	P11512 sulfolobus
23	31	63.3	880	1	RPA1_SULAC	P25655 saccharomyc
24	31	63.3	2108	1	NOT1_YEAST	P19158 saccharomyc
25	31	63.3	3079	1	IRA2_YEAST	Q95276 sus scrofa
26	30	61.2	124	1	RL5_PIG	Q57673 methanococc
27	30	61.2	206	1	ATPE_METJA	O35000 bacillus su
28	30	61.2	242	1	NAGB_BACSU	P49405 caenorhabdi
29	30	61.2	293	1	RL5_CAEEL	P15125 xenopus lae
30	30	61.2	295	1	RL5A_XENLA	P15126 xenopus lae
31	30	61.2	295	1	RL5B_XENLA	P22451 gallus gall
32	30	61.2	296	1	RL5_CHICK	P46777 homo sapien
33	30	61.2	296	1	RL5_HUMAN	P09895 rattus norv
34	30	61.2	296	1	RL5_RAT	
35	30	61.2	301	1	RL5_NEUCR	O59953 neurospora
36	30	61.2	314	1	TRUB_ECOLI	P09171 escherichia
37	30	61.2	387	1	YCBA_BACSU	P40758 bacillus su
38	30	61.2	518	1	HEMK_RICPR	Q9zcb3 rickettsia
39	30	61.2	584	1	65KD_ZYMMO	P15255 zymomonas m
40	30	61.2	592	1	LAM2_MOUSE	P21619 mus musculu
41	30	61.2	604	1	SP20_YEAST	P50075 saccharomyc
42	30	61.2	640	1	YIF6_YEAST	P40522 saccharomyc
43	30	61.2	646	1	SKB1_SCHPO	P78963 schizosacch
44	30	61.2	746	1	TAGF_BACSU	P13485 bacillus su
45	30	61.2	808	1	MCM3_HUMAN	P25205 homo sapien

ALIGNMENTS

RESULT 1
VGLC_HSVE4
ID   VGLC_HSVE4   STANDARD;   PRT;   485 AA.
AC   P22596;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCOPROTEIN C PRECURSOR (GLYCOPROTEIN 13).
GN   GC OR GP13.
OS   Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus
OS   type 1 subtype 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91021040.
RA   NICOLSON L., ONIONS D.E.;
RT   "The nucleotide sequence of the equine herpesvirus 4 gC gene
RT   homologue.";
RL   Virology 179:378-387(1990).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; M58031; AAA46083.1; -.
DR   EMBL; A21044; CAA01528.1; -.
DR   PIR; B45343; B45343.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL   1   32
FT   CHAIN   33   485   GLYCOPROTEIN C.
FT   DOMAIN   33   444   EXTRACELLULAR.
FT   TRANSMEM   445   468
FT   CARBOHYD   60   60   POTENTIAL.
FT   CARBOHYD   61   61   POTENTIAL.
FT   CARBOHYD   66   66   POTENTIAL.
FT   CARBOHYD   67   67   POTENTIAL.
FT   CARBOHYD   72   72   POTENTIAL.
FT   CARBOHYD   108   108   POTENTIAL.
FT   CARBOHYD   116   116   POTENTIAL.
FT   CARBOHYD   147   147   POTENTIAL.
FT   CARBOHYD   220   220   POTENTIAL.
FT   CARBOHYD   225   225   POTENTIAL.
FT   CARBOHYD   286   286   POTENTIAL.
SQ   SEQUENCE   485 AA; 52509 MW; 63F72464 CRC32;

Query Match   69.4%;   Score 34;   DB 1;   Length 485;
Best Local Similarity   70.0%;   Pred. No. 11;
Matches   7;   Conservative   1;   Mismatches   2;   Indels   0;   Gaps   0;

```
Qy      1 YRLLIRLNER 10
           ||||||:|:|
Db    120 YRLEIHLNQR 129

RESULT 2
RL5_BOMMO
ID   RL5_BOMMO      STANDARD;      PRT;    299 AA.
AC   O76190;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L5.
GN   RPL5.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SILK GLAND;
RA   YANG C.S., SEHNAL F.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; AF008229; AAC24960.1; -.
DR   PFAM; PF00861; Ribosomal_L18p; 1.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   299 AA;  34378 MW;  7262D2FC CRC32;

Query Match             67.3%;  Score 33;  DB 1;  Length 299;
Best Local Similarity   50.0%;  Pred. No. 10;
Matches     5;  Conservative   4;  Mismatches   4;  Indels   1;  Gaps   0;

Qy      1 YRLLIRLNER 10
           ||:|:||:.|
Db     49 YRLIVRLSNK 58

RESULT 3
CYL2_HUMAN
ID   CYL2_HUMAN     STANDARD;      PRT;    348 AA.
AC   Q14093;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYLICIN II (MULTIPLE-BAND POLYPEPTIDE II).
GN   CYL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 95255491.
RA   HESS H., HEID H., ZIMBELMANN R., FRANKE W.W.;
RT   "The protein complexity of the cytoskeleton of bovine and human sperm
RT   heads: the identification and characterization of cylicin II.";
RL   Exp. Cell Res. 218:174-182(1995).
CC   -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC       BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; Z46788; CAA86752.1; -.
KW   Cytoskeleton; structural protein; Repeat; Sperm; Spermatogenesis.
FT   DOMAIN       25    347   31 X 3 AA REPEATS OF K-K-X.
FT   DOMAIN      157    240    3 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT      157    184    1.
FT   REPEAT      185    212    2.
FT   REPEAT      213    240    3.
SQ   SEQUENCE   348 AA;  39079 MW;  FD27FBEF CRC32;

Query Match             67.3%;  Score 33;  DB 1;  Length 348;
Best Local Similarity   60.0%;  Pred. No. 12;
Matches     6;  Conservative   3;  Mismatches   3;  Indels   1;  Gaps   0;

Qy      1 YRLLIRLNER 10
           |||||:|:||
Db     78 YRSLMRISER 87

RESULT 4
RFP_HUMAN
ID   RFP_HUMAN      STANDARD;      PRT;    513 AA.
AC   P14373;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ZINC-FINGER PROTEIN RFP (RET FINGER PROTEIN).
GN   RFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88246464.
RA   TAKAHASHI M., INAGUMA Y., HIAI H., HIROSE F.;
RT   "Developmentally regulated expression of a human 'finger'-containing
RT   gene encoded by the 5' half of the ret transforming gene.";
RL   Mol. Cell. Biol. 8:1853-1856(1988).
CC   -!- FUNCTION: MAY FUNCTION IN MALE GERM CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- DISEASE: RECOMBINATION OF RFP WITH A PROTEIN
CC       TYROSINE KINASE PRODUCES THE RET TRANSFORMING PROTEIN.
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; J03407; AAA35564.1; -.
DR   PIR; A28101; TVHURF.
DR   MIM; 602165; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
DR   PFAM; PF00622; SPRY; 1.
DR   PFAM; PF00643; zf-B_box; 1.
KW   Proto-oncogene; Zinc-finger; Metal-binding; Chromosomal translocation;
KW   Nuclear protein; DNA-binding.
FT   SITE        315    316   BREAKPOINT FOR TRANSLOCATION TO FORM THE
FT                           RFP-RET ONCOGENE.
FT   ZN_FING      16     56   C3HC4-TYPE.
```

FT   DOMAIN   96   127   B BOX.
SQ   SEQUENCE   513 AA; 58489 MW; 022BC859 CRC32;

Query Match   67.3%; Score 33; DB 1; Length 513;
Best Local Similarity   77.8%; Pred. No. 18;
Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
          ||||| || |
Db   198 YRLLARLEE 206

RESULT 5
RFP_MOUSE
ID   RFP_MOUSE   STANDARD;   PRT;   522 AA.
AC   Q62158;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ZINC-FINGER PROTEIN RFP (RET FINGER PROTEIN).
GN   RFP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 97176437.
RA   CAO T., SHANNON M., HANDEL M.A., ETKIN L.D.;
RT   "Mouse ret finger protein (rfp) proto-oncogene is expressed at
RT   specific stages of mouse spermatogenesis.";
RL   Dev. Genet. 19:309-320(1996).
CC   -!- FUNCTION: MAY FUNCTION IN MALE GERM CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L46855; AAA85354.1; -.
DR   MGD; MGI:97904; RFP.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
DR   PFAM; PF00622; SPRY; 1.
DR   PFAM; PF00643; zf-B_box; 1.
KW   Zinc-finger; Metal-binding; Nuclear protein; DNA-binding.
FT   ZN_FING   25   65   C3HC4-TYPE.
FT   DOMAIN   105   136   B BOX.
SQ   SEQUENCE   522 AA; 59550 MW; 18E6E716 CRC32;

Query Match   67.3%; Score 33; DB 1; Length 522;
Best Local Similarity   77.8%; Pred. No. 19;
Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
          ||||| || |
Db   207 YRLLARLEE 215

RESULT 6
CHD1_HUMAN
ID   CHD1_HUMAN   STANDARD;   PRT;   1709 AA.
AC   O14646;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)

DE   CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1 (CHD-1).
DE   CHD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97470991.
RA   WOODAGE T., BASRAI M.A., BAXEVANIS A.D., HIETER P., COLLINS F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC   IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF006513; AAB87381.1; -.
DR   HSSP; P23197; 1AP0.
DR   MIM; 602118; -.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00385; chromo; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN   1   70   SER-RICH.
FT   DOMAIN   117   137   SER-RICH.
FT   DOMAIN   272   364   CHROMO DOMAIN.
FT   DOMAIN   389   452   CHROMO DOMAIN.
FT   NP_BIND   506   513   ATP (POTENTIAL).
FT   SITE   614   617   DEAH BOX.
FT   DOMAIN   1628   1644   3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT   1628   1632   1.
FT   REPEAT   1634   1638   2.
FT   REPEAT   1640   1644   3.
SQ   SEQUENCE 1709 AA; 196517 MW; EC7F932A CRC32;

Query Match   67.3%; Score 33; DB 1; Length 1709;
Best Local Similarity   77.8%; Pred. No. 69;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   2 RLLIRLNER 10
          |||||| ||
Db   795 KLLIRLRER 803

RESULT 7
CHD1_MOUSE
ID   CHD1_MOUSE   STANDARD;   PRT;   1711 AA.
AC   P40201;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1 (CHD-1).
DE   CHD1 OR CHD-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 9321972.
RA   DELMAS V., STOKES D.G., PERRY R.P.;
RT   "A mammalian DNA-binding protein that contains a chromodomain and an

```
RT   SNF2/SWI2-like helicase domain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC       IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: ABUNDANCE IS HIGHER IN CELLS REPRESENTING
CC       EARLY STAGES OF THE B LYMPHOID LINEAGE SUCH AS PRE-B AND B CELLS,
CC       THAN IN CELLS REPRESENTING MATURE PLASMACYTES OR OTHER CELL
CC       LINEAGES SUCH AS FIBROBLASTS.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   --------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the  Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------
DR   EMBL; L10410; AAB08486.1; -.
DR   PIR; A47392; A47392.
DR   HSSP; P23197; 1AP0.
DR   MGD; MGI:88393; CHD1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00385; chromo; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN         1     70     SER-RICH.
FT   DOMAIN       116    136     SER-RICH.
FT   DOMAIN       270    362     CHROMO DOMAIN.
FT   DOMAIN       387    450     CHROMO DOMAIN.
FT   NP_BIND      504    511     ATP (POTENTIAL).
FT   SITE         612    615     DEAH BOX.
FT   DOMAIN      1629   1645     3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT      1629   1633     1.
FT   REPEAT      1635   1639     2.
FT   REPEAT      1641   1645     3.
SQ   SEQUENCE   1711 AA;  196409 MW;  CB184D33 CRC32;

Query Match             67.3%;  Score 33;  DB 1;  Length 1711;
Best Local Similarity   77.8%;  Pred. No. 69;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 RLLIRLNER 10
          :||||| ||
Db    793 KLLIRLRER 801

RESULT 8
Y487_MYCTU
ID   Y487_MYCTU    STANDARD;     PRT;    183 AA.
AC   Q11153;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 20.7 KD PROTEIN RV0487.
GN   RV0487 OR MTCY20G9.13.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
```

```
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: STRONG, TO M.LEPRAE U2168E.
CC   --------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the  Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------
DR   EMBL; Z77162; CAB00948.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   183 AA;  20716 MW;  23C141D2 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 183;
Best Local Similarity   60.0%;  Pred. No. 9.7;
Matches    6;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YRLLIRLNER 10
          || |: |||
Db     77 YRFLLRRNRR 86

RESULT 9
YFH8_YEAST
ID   YFH8_YEAST    STANDARD;     PRT;    221 AA.
AC   P43592;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 25.9 KD PROTEIN IN MPR1-GCN20 INTERGENIC REGION.
GN   YFR008W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 95400292.
RA   MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
RA   SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA   YAMAZAKI M., TASHIRO H., EKI T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
CC   --------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the  Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------
DR   EMBL; D50617; BAA09247.1; -.
KW   Hypothetical protein.
FT   DOMAIN     11     22     POLY-GLN.
SQ   SEQUENCE   221 AA;  25915 MW;  3D34764A CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 221;
Best Local Similarity   75.0%;  Pred. No. 12;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YRLLIRLN 8
          ||||:||:|
```

```
Db    211 YRLLLRLH 218

RESULT 10
YJ9D_YEAST
ID   YJ9D_YEAST     STANDARD;      PRT;    239 AA.
AC   P47167;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 27.6 KD PROTEIN IN NMD5-HOM6 INTERGENIC REGION.
GN   YJR135C OR J2122.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROSE M., KOETTER P., ENTIAN K.D.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; Z49635; CAA89666.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   239 AA;  27567 MW;  0BF23C6E CRC32;

Query Match
Best Local Similarity   65.3%;    Score 32;  DB 1;  Length 239;
Matches   6;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     1 YRLLIRLN 8
         ||||:| |
Db   205 YRLLLRAN 212

RESULT 11
DPM1_YEAST
ID   DPM1_YEAST     STANDARD;      PRT;    267 AA.
AC   P14020;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2.4.1.83) (DOLICHOL-
DE   PHOSPHATE MANNOSE SYNTHASE) (DOLICHYL-PHOSPHATE BETA-D-
DE   MANNOSYLTRANSFERASE).
GN   DPM1 OR SED3 OR YPR183W OR P9705.3.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89034276.
RA   ORLEAN P., ALBRIGHT C., ROBBINS P.W.;
RT   "Cloning and sequencing of the yeast gene for dolichol phosphate
RT   mannose synthase, an essential protein.";
RL   J. Biol. Chem. 263:17499-17507(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA   FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,
RA   KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
RA   JOHNSON D., JOHNSTON M., LANGSTON Y., LATREILLE P., LE T.,
RA   MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA   RIFKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
RA   WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE SUGAR DONOR DOL-P-MAN
CC       WHICH IS REQUIRED IN THE SYNTHESIS OF N-LINKED AND O-LINKED
CC       OLIGOSACCHARIDES AND FOR THAT OF GPI ANCHORS.
CC   -!- CATALYTIC ACTIVITY: GDP-MANNOSE + DOLICHYL PHOSPHATE = GDP +
CC       DOLICHYL D-MANNOSYL PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: IF THE N-TERMINUS IS A FUNCTIONAL SIGNAL
CC       SEQUENCE, THE PROTEIN IS PREDICTED TO BE ORIENTED TOWARD THE LUMEN
CC       OF THE ENDOPLASMIC RETICULUM WITH BOTH TERMINI SERVING AS ANCHORS.
CC       THE LACK OF A SIGNAL SEQUENCE INDICATES THAT THE ENZYME FACES THE
CC       CYTOPLASM AND IS ANCHORED AT THE C-TERMINUS.
CC   -!- DOMAIN: THE N-TERMINUS OF THE PROTEIN, THOUGH NOT HYDROPHOBIC,
CC       MEETS EXISTING CRITERIA FOR YEAST SIGNAL SEQUENCES, EVEN THOUGH NO
CC       SITE EXISTS FOR CLEAVAGE BY SIGNAL PEPTIDASE.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------
DR   EMBL; J04184; AAA34578.1; -.
DR   EMBL; U25842; AAB68116.1; -.
DR   PIR; A32122; A32122.
DR   SGD; L0000524; DPM1.
DR   PFAM; PF00535; Glycos_transf_2; 1.
KW   Transferase; Glycosyltransferase; Transmembrane;
KW   Endoplasmic reticulum.
FT   TRANSMEM    239    259       POTENTIAL.
SQ   SEQUENCE   267 AA;  30362 MW;  F8D92784 CRC32;

Query Match
Best Local Similarity   65.3%;    Score 32;  DB 1;  Length 267;
Matches   5;  Conservative   3;  Mismatches   3;  Indels   0;  Gaps   0;

Qy     2 RLLIRLNER 10
         |:::| |||
Db    65 RIIVRTNER 73

RESULT 12
RL5_ANOGA
ID   RL5_ANOGA      STANDARD;      PRT;    327 AA.
AC   O44248;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L5.
GN   RPL5.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC   Culicidae; Anopheles.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G3;
RA   CORNEL A.J., KUMAR V., MUKABAYIRE O., SALAZAR RAFFERTY C.,
RA   PETRARCA V., COLUZZI M., COLLINS F.H.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC   -------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

CC --
DR EMBL; AF002238; AAB97731.1; -.
DR PFAM; PF00861; Ribosomal_L18p; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 327 AA; 37996 MW; F3A3ED2 CRC32;

Query Match
Best Local Similarity 65.3%;   Score 32; DB 1; Length 327;
                       50.0%;   Pred. No. 18;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
         :||::||:|:|
Db   49 FRLIVRLSNR 58

RESULT 13
SAPD_ECOLI
ID SAPD_ECOLI   STANDARD;   PRT;   330 AA.
AC P36635;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PEPTIDE TRANSPORT SYSTEM ATP-BINDING PROTEIN SAPD.
GN SAPD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA EPSTEIN W., NOELKER E., STUMPE S., TEWES R., SCHMID R., BAKKER E.P.;
RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 97251357.
RA AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA ITOH T., KASAI H., KASHIMOTO K., KIMURA S., KITAKAWA M.,
RA KITAGAWA M., MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T.,
RA MOTOMURA K., NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y.,
RA OSHIMA T., SAITO N., SAMPEI G., SEKI Y., SIVASUNDARAM S.,
RA TAGAMI H., TAKEDA J., TAKEMOTO K., TAKEUCHI Y., WADA C.,
RA YAMAMOTO Y., HORIUCHI T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [4]
RP SEQUENCE OF 301-330 FROM N.A.
RA BERGLER H., EBELING A., FUCHSBICHLER S., HOGENAUER G.,
RA TURNOWSKY F.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: INVOLVED IN A PEPTIDE INTAKE TRANSPORT SYSTEM THAT
CC   PLAYS A ROLE IN THE RESISTANCE TO ANTIMICROBIAL PEPTIDES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC   (ABC TRANSPORTERS).
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X97282; CAA65940.1; -.
DR EMBL; AE000227; AAC74373.1; -.
DR EMBL; D90766; CAB20760.1; -.
DR EMBL; D90767; CAB20768.1; -.
DR EMBL; D90768; CAB20789.1; -.
DR EMBL; U08190; AAA17670.1; -.
DR ECOGENE; EG12304; SAPD.
DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR PFAM; PF00005; ABC_tran; 1.
KW Peptide transport; Transport; Inner membrane; ATP-binding.
FT NP_BIND   40   47   ATP (POTENTIAL).
SQ SEQUENCE 330 AA; 37660 MW; BE0692E6 CRC32;

Query Match
Best Local Similarity 65.3%;   Score 32; DB 1; Length 330;
                       66.7%;   Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy   1 YRLLIRLNE 9
         :||| |||:
Db 198 FRLLTRLNQ 206

RESULT 14
SAPD_SALTY
ID SAPD_SALTY   STANDARD;   PRT;   330 AA.
AC P36636;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PEPTIDE TRANSPORT SYSTEM ATP-BINDING PROTEIN SAPD.
GN SAPD.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14028S;
RX MEDLINE; 94038887.
RA PARRA-LOPEZ C., BAER M.T., GROISMAN E.A.;
RT "Molecular genetic analysis of a locus required for resistance to
RT antimicrobial peptides in Salmonella typhimurium.";
RL EMBO J. 12:4053-4062(1993).
CC -!- FUNCTION: INVOLVED IN A PEPTIDE INTAKE TRANSPORT SYSTEM THAT
CC   PLAYS A ROLE IN THE RESISTANCE TO ANTIMICROBIAL PEPTIDES.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC   (ABC TRANSPORTERS).
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch).
CC --
DR EMBL; X74212; CAA52287.1; -.
DR PIR; S39588; S39588.
DR STYGENE; SG10380; SAPD.
DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR PFAM; PF00005; ABC_tran; 1.
KW Peptide transport; Transport; Inner membrane; ATP-binding.
FT NP_BIND   40   47   ATP (POTENTIAL).
SQ SEQUENCE 330 AA; 37611 MW; 9910EB90 CRC32;

Query Match
Best Local Similarity 65.3%;   Score 32; DB 1; Length 330;
                       66.7%;   Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 YRLLIRINE 9
           :||| |||:
Db     198 FRLLTRLNQ 206

RESULT 15
VGLC_HSVEB
ID   VGLC_HSVEB     STANDARD;     PRT;    468 AA.
AC   P12889; P36321;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCOPROTEIN C PRECURSOR (GLYCOPROTEIN 13).
GN   GC OR GP13 OR 16.
OS   Equine herpesvirus type 1 (strain Ab4p) (EHV-1), and
OS   Equine herpesvirus type 1 (strain Kentucky D) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB4P;
RX   MEDLINE; 92295566.
RA   TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVISON A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KENTUCKY D;
RX   MEDLINE; 88275055.
RA   ALLEN G.P., COOGLE L.D.;
RT   "Characterization of an equine herpesvirus type 1 gene encoding a
RT   glycoprotein (gp13) with homology to herpes simplex virus
RT   glycoprotein C.";
RL   J. Virol. 62:2850-2858(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KENTUCKY D;
RX   MEDLINE; 89382761.
RA   GUO P., GOEBEL S., DAVIS S., PERKUS M.E., LANGUET B., DESMETTRE P.,
RA   ALLEN G., PAOLETTI E.;
RT   "Expression in recombinant vaccinia virus of the equine herpesvirus 1
RT   gene encoding glycoprotein gp13 and protection of immunized
RT   animals.";
RL   J. Virol. 63:4189-4198(1989).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KENTUCKY D;
RX   MEDLINE; 93212524.
RA   MATSUMURA T., SMITH R.H., O'CALLAGHAN D.J.;
RT   "DNA sequence and transcriptional analyses of the region of the
RT   equine herpesvirus type 1 Kentucky A strain genome encoding
RT   glycoprotein C.";
RL   Virology 193:910-923(1993).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L07272; AAA46078.1; -.
DR   EMBL; M86664; AAB02451.1; -.
DR   EMBL; M19966; AAA46077.1; -.
DR   EMBL; M29234; AAA46085.1; -.
DR   EMBL; S57839; AAB25944.1; -.
DR   PIR; H36796; VGBEA1.
DR   PIR; A28149; VGBEEH.
DR   PIR; B46114; B46114.
```

```
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    468       GLYCOPROTEIN C.
FT   TRANSMEM    432    451       POTENTIAL.
FT   CARBOHYD     46     46       POTENTIAL.
FT   CARBOHYD     57     57       POTENTIAL.
FT   CARBOHYD     62     62       POTENTIAL.
FT   CARBOHYD     92     92       POTENTIAL.
FT   CARBOHYD    100    100       POTENTIAL.
FT   CARBOHYD    131    131       POTENTIAL.
FT   CARBOHYD    203    203       POTENTIAL.
FT   CARBOHYD    208    208       POTENTIAL.
FT   CARBOHYD    269    269       POTENTIAL.
FT   CONFLICT    107    107       E -> K (IN REF. 4).
FT   CONFLICT    145    145       E -> K (IN REF. 4).
FT   CONFLICT    275    275       V -> A (IN REF. 4).
SQ   SEQUENCE   468 AA;  50889 MW;  EBF20B7 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 468;
Best Local Similarity   70.0%;
Matches    7;  Conservative   1;  Mismatches   2;  Indels   0;  Gaps   0;

Qy        1 YRLLIRLNER 10
            |||:| ||:|
Db      104 YRLEIYLNQR 113


Search completed: February  8, 2000, 00:59:52
Job time: 3781 sec
```

OM protein - protein search, using sw model

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: February 8, 2000, 13:17:37 ; Search time 209.03 Seconds
            (without alignments)
            3.317 Million cell updates/sec

Title: US-08-653-294-11
Perfect score: 49
Sequence: 1 YRLLIRLNER 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08
         Listing first 45 summaries

Database :   SPTREMBL_12:*
```
 1: sp_archea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	73.5	395	10	Q9ZT98	q9zt98 arabidopsis
2	34	69.4	254	11	P70561	p70561 rattus norv
3	34	69.4	485	12	O39258	o39258 equine herp
4	33	67.3	191	13	Q98543	q98543 paramecium
5	33	67.3	201	3	Q12530	q12530 saccharomyc
6	33	67.3	506	11	O62157	o62157 mus musculu
7	33	67.3	515	2	Q9ZAP8	q9zap8 paenibacill
8	33	67.3	537	5	Q22520	q22520 caenorhabdi
9	33	67.3	560	13	Q08781	q08781 gallus gall
10	33	67.3	859	11	P70261	p70261 mus musculu
11	33	67.3	1808	13	O42142	o42142 gallus gall
12	32	65.3	151	1	O58407	o58407 pyrococcus
13	32	65.3	261	5	O96646	o96646 drosophila
14	32	65.3	678	5	P71707	p71707 mycobacteri
15	31	63.3	214	10	O65654	o65654 arabidopsis
16	31	63.3	260	5	O76669	o76669 caenorhabdi
17	31	63.3	329	2	O86190	o86190 erwinia chr
18	31	63.3	397	10	O48758	o48758 arabidopsis
19	31	63.3	418	2	Q9X8L1	q9x8l1 streptomyce
20	31	63.3	425	5	Q19125	q19125 caenorhabdi
21	31	63.3	442	8	Q9XMT9	tetrahymena
22	31	63.3	570	8	O13347	magnaporthe
23	31	63.3	688	3	Q34312	dictyostell
24	31	63.3	819	3	O94254	schizosacch
25	31	63.3	1097	3	O13592	saccharomy
26	31	63.3	1131	10	O23741	brassica ol
27	30	61.2	81	12	Q9W179	human immun
28	30	61.2	144	2	Q27369	trypanosoma
29	30	61.2	173	5	Q51316	nostoc sp.
30	30	61.2	175	5	Q17265	brugia paha
31	30	61.2	176	2	O32098	bacillus su
32	30	61.2	186	2	O67378	aquifex aeo
33	30	61.2	189	12	Q98542	paramecium
34	30	61.2	216	2	O85823	yersinia ps
35	30	61.2	296	5	Q9Y0H6	myxine glut
36	30	61.2	309	13	Q91634	xenopus lae
37	30	61.2	366	2	Q9ZBT9	streptomyce
38	30	61.2	367	10	O80931	arabidopsis
39	30	61.2	378	2	O87101	bacillus su
40	30	61.2	394	8	Q34835	kluyveromyc
41	30	61.2	431	13	Q9W734	gallus gall
42	30	61.2	437	5	Q19360	caenorhabdi
43	30	61.2	447	5	Q21147	caenorhabdi
44	30	61.2	460	2	P94248	bifidobacte
45	30	61.2	471	5	Q9Y1X6	ephydatia f

ALIGNMENTS

```
RESULT 1
Q9ZT98
ID   Q9ZT98       PRELIMINARY;     PRT;    395 AA.
AC   Q9ZT98;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE LEUCINE-RICH REPEAT PROTEIN.
GN   T419.11.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-CV. COLUMBIA;
RA   PARNELL L.D., GNOJ L., DE LA BASTIDE M., HAMEED A., HABERMANN K.,
RA   SCHUTZ K., HUANG E., GOTTESMAN T., DEDHIA N.N., MCCOMBIE W.R.,
RT   "Genomic sequence of BAC T419 from Arabidopsis thaliana, Chromosome
RT   IV, near 16.6 CM.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF069442; AAC79105.1; -.
SQ   SEQUENCE   395 AA;  43507 MW;  F547BBD5 CRC32;

Query Match             73.5%;  Score 36;  DB 10;  Length 395;
Best Local Similarity   66.7%;  Pred. No. 15;
Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YRLLIRLNE 9
           ||:|:|||:
Db     231 YRVLLRLNQ 239


RESULT 2
P70561
ID   P70561       PRELIMINARY;     PRT;    254 AA.
AC   P70561;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   FGF RECEPTOR ACTIVATING PROTEIN FRAG1.
```

```
GN   FRAG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   LORENZI M.V., HORII Y., YAMANAKA R., SAKAGUCHI K., MIKI T.;
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
DR   EMBL; U57715; AAB07050.1; -.
SQ   SEQUENCE  254 AA;  29395 MW;  8470603F CRC32;

Query Match             69.4%;  Score 34;  DB 11;  Length 254;
Best Local Similarity   87.5%;  Pred. No. 25;
Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YRLLIRLN 8
           |||| |||
Db     109 YRLLCRLN 116


RESULT 3
039258
ID   039258          PRELIMINARY;          PRT;    485 AA.
AC   039258;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   COUNTERPART OF HSV-1 GENE UL44 AND V2V GENE 14.
GN   16.
OS   Equine herpesvirus 4.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NS80567;
RX   MEDLINE; 91021040.
RA   TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RA   NICOLSON L., ONIONS D.E.;
RT   "The nucleotide sequence of the equine herpesvirus 4 gC gene
RT   homologue.";
RL   Virology 179:378-387(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NS80567;
RX   MEDLINE; 98264497.
RA   TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RT   "The DNA sequence of equine herpesvirus-4.";
RL   J. Gen. Virol. 79:1197-1203(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NS80567;
RA   TELFORD E.A., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF030027; AAC59530.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE  485 AA;  52539 MW;  8DF52A42 CRC32;

Query Match             69.4%;  Score 34;  DB 12;  Length 485;
Best Local Similarity   70.0%;  Pred. No. 48;
Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YRLLIRLNER 10
           ||| ||:||:|
Db     120 YRLEIHLNQR 129


RESULT 4
Q98543
ID   Q98543          PRELIMINARY;          PRT;    191 AA.
AC   Q98543;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GENOME, PARTIAL SEQUENCE.
GN   A493L.
OS   Paramecium bursaria chlorella virus 1 (PBCV-1).
OC   Viruses; dsDNA viruses, no RNA stage: Phycodnaviridae; Phycodnavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96400190.
RA   KUTISH G.F., LI Y., LU Z., FURUTA M., ROCK D.L., VAN ETTEN J.L.;
RT   "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
RT   positions 182 to 258.";
RL   Virology 223:303-317(1996).
DR   EMBL; U42580; AAC96860.1; -.
SQ   SEQUENCE  191 AA;  22651 MW;  E4547C83 CRC32;

Query Match             67.3%;  Score 33;  DB 12;  Length 191;
Best Local Similarity   66.7%;  Pred. No. 30;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps    0;

Qy       1 YRLLIRLNE 9
           |||:||:| ||
Db      74 YRLVVRENE 82


RESULT 5
Q12530
ID   Q12530          PRELIMINARY;          PRT;    201 AA.
AC   Q12530;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   CHROMOSOME XII READING FRAME ORF YLR145W.
GN   L9634.3.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RIEGER M., MUELLER-AUER S., BRUECKNER M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN-S288C (AB972);
RA   FULTON L.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN-S288C (AB972);
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA   FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
RA   HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
RA   JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
RA   MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN D., RILES L.,
RA   TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,
RA   WILSON R., WATERSTON R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN-S288C (AB972);
RA   WATERSTON R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73317; CAA97717.1; -.
DR   EMBL; U53879; AAB82379.1; -.
SQ   SEQUENCE  201 AA;  23618 MW;  FDD081D6 CRC32;

Query Match             67.3%;  Score 33;  DB 3;  Length 201;
```

Best Local Similarity 60.0%;   Pred. NO. 32;
Matches   6; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNER 10
             ||||:: || |
Db   15 YRLLILLNHR 24

RESULT   6
ID   Q62157   PRELIMINARY;   PRT;   506 AA.
AC   Q62157;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ZINC FINGER PROTEIN (FRAGMENT).
GN   RFP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RC   STRAIN=BALB/C;
RA   TAKAHASHI M.;
RL   Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR   EMBL; X75343; CAA53092.1; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00622; SPRY; 1.
DR   PFAM; PF00643; zf-B_box; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
KW   DNA-binding; Zinc-finger.
FT   NON_TER   1   1
SQ   SEQUENCE   506 AA;   57882 MW;   AEE397C3 CRC32;

Query Match   67.3%;   Score 33; DB 11; Length 506;
Best Local Similarity   77.8%;   Pred. No. 79;
Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
             |||| || |
Db   191 YRLLARLEE 199

RESULT   7
ID   Q9ZAP8   PRELIMINARY;   PRT;   515 AA.
AC   Q9ZAP8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEOPULLULANASE.
OS   Paenibacillus polymyxa (Bacillus polymyxa).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Paenibacillus.
RN   [1]
RC   STRAIN=CECT 155;
RX   MEDLINE; 99118304.
RA   YEBRA M.J., BLASCO A., SANZ P.;
RT   "Expression and secretion of Bacillus polymyxa neopullulanase in
RT   Saccharomyces cerevisiae.";
RL   FEMS Microbiol. Lett. 170:41-49(1999).
DR   EMBL; U89716; AAD05199.1; -.
DR   HSSP; P21332; 1UOK.
SQ   SEQUENCE   515 AA;   58749 MW;   68075B21 CRC32;

Query Match   67.3%;   Score 33; DB 2; Length 515;
Best Local Similarity   77.8%;   Pred. No. 81;
Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLNE 9
             |||||||| |
Db   419 YRTLIRLRE 427

RESULT   8
ID   Q22520   PRELIMINARY;   PRT;   537 AA.
AC   Q22520;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T16A9.1 PROTEIN.
GN   T16A9.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RA   MCMURRAY A.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z77135; CAB00875.1; -.
SQ   SEQUENCE   537 AA;   61447 MW;   A7597F26 CRC32;

Query Match   67.3%;   Score 33; DB 5; Length 537;
Best Local Similarity   55.6%;   Pred. No. 84;
Matches   5; Conservative   4; Mismatches   0; Indels   0; Gaps   0;

Qy   2 RLLIRLNER 10
             :|:|::||||
Db   113 KLIVKLNER 121

RESULT   9
ID   Q08781   PRELIMINARY;   PRT;   560 AA.
AC   Q08781;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 64.4 KD PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LENS;
RX   MEDLINE; 94116444.
RA   FUNAHASHI J., SEKIDO R., MURAI K., KAMACHI Y., KONDOH H.;
RT   "Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-
RT   homeodomain protein implicated in postgastrulation embryogenesis.";
RL   Development 119:433-446(1993).
DR   EMBL; D14316; BAA03262.1; -.
DR   PFAM; PF00271; helicase_C; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   560 AA;   64394 MW;   A54C9E16 CRC32;

Qy   -2 RLLIRLNER 10
         :||||| ||
Db   43 KLLIRLRER 51

RESULT 10
P70261
ID   P70261   PRELIMINARY;   PRT;   859 AA.
AC   P70261;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PALADIN GENE.
GN   PALADIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   PEARCE J.J.H., DAVIES T., GARDENER R.L.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X99384; CAA67763.1; -.
SQ   SEQUENCE   859 AA; 96739 MW; 8D061D00 CRC32;

Query Match   67.3%; Score 33; DB 13; Length 560;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches   6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

Qy   1 YRLLIRLN 8
        ||||: ||
Db 464 YRLLVELN 471

RESULT 11
O42142
ID   O42142   PRELIMINARY;   PRT;   1808 AA.
AC   O42142;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CHROMO-HELICASE-DNA-BINDING ON THE 2 CHROMOSOME PROTEIN.
GN   CHD-2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97473516.
RA   GRIFFITHS R., KORN R.M.;
RT   "A CHD1 gene is 2 chromosome linked in the chicken Gallus
RT   domesticus.";
RL   Gene 197:225-229(1997).
DR   EMBL; AF004397; AAC60282.1; -.
DR   HSSP; P23197; 1AP0.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PFAM; PF00385; chromo; 2.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00176; SNF2_N; 1.
KW   Helicase; DNA-binding.
SQ   SEQUENCE   1808 AA; 208399 MW; 97FEB926 CRC32;

Query Match   67.3%; Score 33; DB 13; Length 1808;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches   7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy   2 RLLIRLNER 10
         :||||| ||
Db   792 KLLIRLRER 800

RESULT 12
O58407
ID   O58407   PRELIMINARY;   PRT;   151 AA.
AC   O58407;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   151AA LONG HYPOTHETICAL FRXA PROTEIN.
GN   PH0674.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000003; BAA29765.1; -.
SQ   SEQUENCE   151 AA; 17160 MW; 11AACD59 CRC32;

Query Match   65.3%; Score 32; DB 1; Length 151;
Best Local Similarity 77.8%; Pred. No. 39;
Matches   7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy   2 RLLIRLNER 10
         |||| |:||
Db   96 RLLIELDER 104

RESULT 13
O96646
ID   O96646   PRELIMINARY;   PRT;   261 AA.
AC   O96646;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PIGMENT CELL DEHYDROGENASE REDUCTASE.
GN   PCDR.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98384483.
RA   BRUNEL C.A., MADIGAN S.J., CASSILL J.A., EDEEN P.T., MCKEOWN M.;
RT   "pcdr, a novel gene with sexually dimorphic expression in the pigment
RT   cells of the Drosophila eye.";
RL   Dev. Genes Evol. 208:327-335(1998).
DR   EMBL; AF098864; AAC72391.1; -.
SQ   SEQUENCE   261 AA; 28302 MW; DC731F30 CRC32;

Query Match   65.3%; Score 32; DB 5; Length 261;
Best Local Similarity 60.0%; Pred. No. 66;
Matches   6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy   1 YRLLIRLNER 10
        ||:| ||:::
Db 207 YRILDRLNKQ 216

```
RESULT 14
P71707
ID   P71707         PRELIMINARY;      PRT;     678 AA.
AC   P71707;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROBABLE PENICILLIN-BINDING PROTEINS 1A/1B (PBP1).
GN   PONA OR RV0050 OR MTCY21.13.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E., TEKAIA F., BADCOCK K.,
RA   BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R., DAVIES R.,
RA   DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S., HORNSBY T.,
RA   JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L., OLIVER S.,
RA   OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J., RUTTER S.,
RA   SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E., TAYLOR K.,
RA   WHITEHEAD S., BARRELL B.G.;
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: CELL WALL FORMATION.
CC   -!- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER BACTERIAL CLASS 1A PENICILLIN-BINDING
CC       PROTEINS.
DR   EMBL; Z80075; CAB02529.1; -.
DR   PFAM; PF00912; Transglycosyl; 1.
DR   PFAM; PF00905; Transpeptidase; 1.
KW   Peptidoglycan synthesis; Cell wall; Transmembrane.
FT   ACT_SITE     347    347       ACYLATED BY PENICILLIN (BY SIMILARITY).
FT   DOMAIN       625    631       POLY-PRO.
FT   DOMAIN       664    670       POLY-PRO.
SQ   SEQUENCE   678 AA;  71150 MW;  58CC83F1 CRC32;

Query Match             65.3%;  Score 32;  DB 2;  Length 678;
Best Local Similarity   62.5%;  Pred. No. 1.7e+02;
Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YRLLIRLN 8
            ||||::||
Db    404 YRLMLKLN 411


RESULT 15
O65654
ID   O65654         PRELIMINARY;      PRT;     214 AA.
AC   O65654;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 24.8 KD PROTEIN.
GN   T19P19.60.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., MONFORT A., CASACUBERTA E., PUIGDOMENECH P., HOHEISEL J.,
RA   MEWES H.W., MAYER K.F.X., SCHUELLER C.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL022605; CAA18753.1; -.
DR   MENDEL; 29179; Arath;3410:29179.
KW   Hypothetical protein.
SQ   SEQUENCE   214 AA;  24813 MW;  110A2C72 CRC32;

Query Match             63.3%;  Score 31;  DB 10;  Length 214;
Best Local Similarity   75.0%;  Pred. No. 87;
Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 RLLIRLNE 9
            :||:||||
Db    171 QLLLRLNE 178


Search completed: February   8, 2000, 13:17:39
Job time: 32488 sec
```

OM of: US-08-653-294-11 to: GenEmbl:*   out_format : pfs
Date: Feb 8, 2000 4:38 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=frame+_p2n.model -DEV=xlp
  -Q=/cgn1/USPTO_spool/US08653294/runat.04022000_160701_15779/app_query.fasta.1
  -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
  -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPEXT=4.500
  -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
  -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
  -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
  -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
  -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
  -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
  Query: US-08-653-294-11
  Query length: 10
  Database: GenEmbl:*
  Database sequences: 821193
  Database length: -1518192014
  Search time (sec): 11370.480000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_pll:AP000815	+	41.00	100.04	1.5e+03	142418	! AP000815 Oryza sativa genomi
gb_in1:CEC50B6	+	40.00	106.62	646.26	41322	! Z81050 Caenorhabditis elegans
gb_htg7:AC017780	-	40.00	103.29	991.28	60818	! AC017780 Drosophila melanogas
gb_htg3:AC009355	-	40.00	102.46	1.1e+03	66958	! AC009355 Drosophila melanogas
gb_htg3:AC005365	+	40.00	100.28	1.5e+03	86130	! AC005365 Homo sapiens chromos
gb_pr3:HS1141E20	-	40.00	99.18	1.7e+03	97906	! AL109912 Homo sapiens chromos
gb_htg3:AC008324	-	40.00	97.27	2.1e+03	122061	! AC008324 Drosophila melanoga
gb_pr3:HS1112D6	+	40.00	96.38	2.4e+03	135305	! AL080317 Human DNA sequence
gb_htg7:AC017912	-	40.00	93.66	3.4e+03	185469	! AC017912 Drosophila melanoga
gb_htg7:AC017912	-	39.00	101.01	1.3e+03	49261	! AC017912 Drosophila melanoga
gb_pr2:AC002465	+	39.00	91.07	4.8e+03	155881	! AC002465 Human BAC clone RG3
gb_htg5:AC015280	-	39.00	90.87	4.9e+03	159468	! AC015280 Drosophila melanoga
gb_htg4:AC007054	-	39.00	90.22	5.3e+03	171979	! AC007054 Drosophila melanoga
gb_htg4:AC006467	+	39.00	90.03	5.4e+03	175695	! AC006467 Drosophila melanoga
gb_htg3:AC008682	-	39.00	88.22	6.5e+03	216649	! AC008682 Homo sapiens chromo
gb_ro:AB012933	-	38.00	122.79	81.19	2454	! AB012933 Rattus norvegicus mRN
gb_pll:ATHSP881	+	38.00	117.48	160.57	4544	! Y11829 A.thaliana hsp88.1 gene
gb_htg3:AC008847	+	38.00	98.65	1.8e+03	40252	! AC008847 Homo sapiens chromos
gb_in2:AC004287	-	38.00	93.77	3.4e+03	70841	! AC004287 Drosophila melanogas
gb_htg4:AC010574	+	38.00	93.26	3.6e+03	75163	! AC010574 Drosophila melanogas
gb_htg7:AC017187	-	38.00	91.37	4.6e+03	93588	! AC017187 Drosophila melanogas
gb_pr4:AC000100	+	38.00	90.64	5.0e+03	101912	! AC000100 Homo sapiens Chromo
gb_pl2:ATAC007167	+	38.00	90.47	5.1e+03	103874	! AC007167 Arabidopsis thalian
gb_htg5:AC009251	-	38.00	89.96	5.5e+03	110262	! AC009251 Drosophila melanoga
gb_htg4:AC010121	+	38.00	89.72	5.6e+03	113253	! AC010121 Drosophila melanoga
gb_htg4:AC010006	+	38.00	89.71	5.7e+03	113508	! AC010006 Drosophila melanoga
gb_htg3:AC010112	-	38.00	89.08	6.1e+03	122073	! AC010112 Drosophila melanoga
gb_htg3:AC010682	+	38.00	87.16	7.8e+03	152393	! AC010682 Homo sapiens clone
gb_pl2:AC013934	+	38.00	86.24	8.8e+03	169516	! AC013934 Drosophila melanoga
gb_htg4:AC009659	+	38.00	85.06	1.0e+04	194446	! AC009659 Homo sapiens chromo
gb_bal:BACCRYMT	-	37.00	120.04	115.57	2100	! M31737 B.thuringiensis mosquit
gb_pat:I09413	+	37.00	120.04	115.57	2100	! I09413 Sequence 1 from Patent
gb_pl2:AF203588	-	37.00	118.38	142.97	2545	! AF203588 Pterocarpus macrocarp
gb_pl2:AF142691	-	37.00	118.36	143.34	2551	! AF142691 Pterocarpus indicus m
gb_pl1:BNAMTNAD3	+	37.00	118.28	144.90	2576	! D13697 Brassica napus mitochon
gb_pl2:AF142728	+	37.00	118.22	145.96	2593	! AF142728 Robinia pseudoacacia
gb_in1:CEC07G2	-	37.00	100.29	1.5e+03	20720	! Z32840 Caenorhabditis elegans
gb_htg5:AC014301	+	37.00	96.98	2.2e+03	30386	! AC014301 Drosophila melanogas
gb_pl2:F19K23	-	37.00	86.92	8.1e+03	97451	! AC007583 Sequence of BAC F19K
gb_htg2:AC007583	+	37.00	86.05	9.0e+03	107818	! AC007583 Arabidopsis thalian
gb_pr3:HSJ365O19	+	37.00	82.81	1.4e+04	156935	! AL096867 Human DNA sequence
gb_htg4:AC011274	-	37.00	82.27	1.5e+04	167074	! AC011274 Homo sapiens clone

gb_htg6:AC013557	+	37.00	81.70	1.6e+04	178506	! AC013557 Homo sapiens chr
gb_htg4:AC011189	+	37.00	81.65	1.6e+04	179611	! AC011189 Homo sapiens chr
gb_htg3:AC008537	-	37.00	81.29	1.7e+04	187246	! AC008537 Homo sapiens chr

seq_name: gb_pll:AP000815

seq_documentation_block:
LOCUS   AP000815   142418 bp DNA   PLN   04-DEC-1999
DEFINITION Oryza sativa genomic DNA, chromosome 1, clone:P0003H10.
ACCESSION   AP000815
VERSION   AP000815.1   GI:6498418
KEYWORDS   .
SOURCE   Oryza sativa (cultivar:Nipponbare) DNA, clone:P0003H10.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 142418)
AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE   Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0003H10
JOURNAL   Published Only in Database (1999) In press
REFERENCE   2 (bases 1 to 142418)
AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL   Direct Submission
            Submitted (30-NOV-1999) to the DDBJ/EMBL/GenBank databases. Takuji
            Sasaki, National Institute of Agrobiological Resources, Rice Genome
            Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, Tel:81-298-38-7441,
            URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441,
            Fax:81-298-38-7468)
COMMENT   The orientation of the sequence is from T7 to SP6 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October 1998 version). The genomic sequence was
            searched against the non-redundant database NRP (PIR, SWISSPROT,
            GENPEPT, PDB) from MAFF DNA bank and the cDNA sequence database at
            RGP. Protein similarities of the cording regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN 2.0 with the corresponding DDBJ accession
            no. and RGP clone ID.
            Detailed information on assemble quality together with annotation
            of this entry at http://www.dna.affrc.go.jp:82/genomicdata/GenomeFi
            nished.html.
FEATURES   Location/Qualifiers
     source   1..142418
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
                     /clone="P0003H10"
     CDS   join(2901..2951,3129..3317,3521..3850,4405..4638,
                     5069..5509)
                     /note="Similar to Arabidopsis thaliana DNA chromosome 4,
                     BAC clone F22K18 (AL035356)"
                     /codon_start=1
                     /protein_id="BAA87822.1"
                     /db_xref="GI:6498419"
                     /translation="MGEPNQRRLYQVWQGSNKFLCGGRLIFGPDAGSLFLSTVLIVAP
                     LVGLCCQCITKMNS1SSEKQVLGLPVLIATIVLGLADLALFLLMTSSRDPGIVPRNARP
                     PESCGGGDEEVAGDVTTPSAEWTAASPHLRLPRSKDVVVNGCVMKVKYCDTCLLYR
                     PPRASHCSINNCVRKFDHICPWVGQCIGLRVNRFFLFISTSTLLCVYFVSWLNI
                     VAHKDGNDGSLLKSMAGEPLSVVLIVYTFVSVWFVGGLTVEHLYLMSTNQTTYENFRY
                     RYDKKENPYNRGAISNIAEVFCAGIPPSMNNFRSWVAPPLEEPDVSGQLPPRNGAD
                     LTGGVKEKVDLEMGRNGGIPAILRGLDYDMEKNDVSVHIKDRGAAPAADPDFMAGR
                     WHNEDCESTPTAVSHHVNSERL"
     CDS   join(9165..9635,9735..9799,9992..10319,10422..10520,
                     10743..11005)
                     /note="ESTs D47452(S12946),C23573(S10086),C19173(E10057),
                     AU081294(E10057),AU032455(S10086) correspond to a region
                     of the predicted gene. Similar to PNIL34. (U37437)"
                     /codon_start=1
                     /protein_id="BAA87823.1"

CDS

```
/db_xref="GI:6498420"
/translation="MASSAMELSLLNPAAMRGLSAAKPRVVSSRRIVRFRVASSAAAP
PAAKPGTPKKRGKTEIQETLLTPRFYTTDFDEMERLFNAEINKQLNQEEFDALLQEFK
TDYNQTHFVRNPEFKAAADKMEGPLRQIFVELERSCTAEFSGFLLYKELGRRLKKTN
PVVAEIFSLMSRDEARHAGFLNKGLSDFNLADLGFVKARKYTFFKPKFIFYATYLS
EKIGYWRYITIFRHLKANPEYQVYPIFKYFENWCQDENRHGDFFSALLKAQPQFLNDW
KAKLWSRFFCLSVYVTMYLNDCQRTTFYEGICLDTKFDMHVIIETNRTTARIFPAVL
DVENPEFKRKLDRMVEINKKIIAIGESDIPLVKNLKRIPHVAALVSEIIRAYLMPPI
ESGSVDFAEFEPQLYV"
join(13373. .13457,15836. .15914,17237. .17477,17560. .17634,
17726. .17860,18400. .18587,18681. .18749,18826. .18916,
19028. .19162,19323. .19379,19441. .19450,19509,
19523. .19585,19589. .19701,19800. .19941,20057. .20294,
21399. .21586)
/note="ESTs AU081256(C53656),AU081257(C53656),
AU032236(R3784),AU081351(E61905),AU031587(E61905),
AU081361(R3784) correspond to a region of the predicted
gene.; Similar to alien-like protein. (AC005623)"
/codon_start=1
/protein_id="BAA87824.1"
/db_xref="GI:6498421"
/translation="MEDYGFEYSDDEPEEQDVDIENQYINSKGMVETDPEGALAGFDQ
VVRMEPEAEWGFKALKQTVKLYYKLGKYKEMDAYREMLIYIKSAVTRNYSEKCINN
IMDFVSGSASQNFSLLQEFYQTLKALEEAKNERLWFKTNLKLCKIWFDMGEYGRMSK
ILKELRKSCQREDGSDDQKKGTQLLEVYAIEIQMYTETKNNKLKELYTKALSIKSAI
PHPRIMGIIRECGKMHMAERQWADAATDFEEAFKNVDEAGNPRRIQCLKIYIVLANML
MESEVNPFDGEAKPYKNDPEILAMTNLIAAYQKNDIMEFKILKSNRRTIMDDPFIR
NYIEDLLKNIRTQVLLKLIKPTRIRIPFISQKWCMKIISNMELWYILVLFHGTLVYT
RAILGSCKKYSLHLDFLSCAQTRVQLKSCALNRADPMLFLTEVNLLSYSVDALSSI
VLSMFCMELNPEKDVEQLLVSLILDNRIQGHIDQVNKLLERGDREIPSHRQVEYSAE
EHLPNGVQQSWREIAGCLGCTALGTCLHVLLWMLGFCPELNLSANSNPEETILASM
ARNFWFCWHCRTSKFWVRGNWFRVKYAPGTRHTANIVPLPSSDSVANQLSNQNG
KESPFLSYTTDQQALNLQREHTTLSDNVINLSEGSYRPLAPSSHLGHPVTTIFYVDDN
I"
join(22475. .22550,23332. .23705))
/note="ESTs AU030013(E50493),AU081341(E50493) correspond
to a region of the predicted gene.; Similar to O.sativa
gene encoding calmodulin. (Z12828)"
/codon_start=1
/protein_id="BAA87825.1"
/db_xref="GI:6498422"
/translation="MADQLTDDQIAEFKEAFSLFDKDGDGCITTKELGTVMRSLGQNP
TEAELQDMINEVDADGNGTIDFPEFLNLMARKMKDTDSEEELKEAFRVFDKDQNGFIS
AAELRHVTNLGEKLTDEEVEMIREADVDGDGQINYEFEVVMMAK"
complement(join(24778. .24897,25615. .25722,25808. .25893,
26090. .26177,26435. .26638,27519. .27581,27708. .27778,
28329. .28389,28812. .28552,29058. .29225,30476. .31136))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87826.1"
/db_xref="GI:6498423"
/translation="MWFQTSTSDARTHGSGDRGSAGRIRACRGTTRSAAPGLGSREAT
TRGHALAEPEPHASPRLVGHRGSRRRSPSPPSPPPLPVTARCEEEHRRHRCSSLFPR
RQMRRGAPPPDPASPCPDLATLGVGRRWKKWDAAEEGVGAEEEERMARGRARQLRGG
GGGVVAHKRRSGAREEERRCCGFERRSGGDERGGVREEEDKAEPVGRRENKGERIGE
TDERRMGHSHKPSIIIMWIQTKPARRGANVLARHCGDDRRRCRSQCAVARRRGE
ALPIQSLFPYLYFMIRDLKVAKEEQDIGFYAGFVASCIVQAYASEVCRKEHQALGISL
VTSSRAIALVVGPAIGGFLSQETLHMHHDDKEVIDALEAQDATSDLGETTKESGSGRM
GHTKSLLKNWQLMSAITLYCVFSLHDTAYLEIFSLNAVSSRKYRGLSFTSQDVGIVLA
ISGFGVLVYQLAIYPLLAKYVGPIKPFRYAAVLSILLSTYPFMANLYGLELKVLINI
ASLLKNMFANKKEMTNQSISLLDAGFHGRKSTLLGCSYQVKQQGSSKP"
complement(join(34468. .34544,37202. .37385))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87827.1"
/db_xref="GI:6498424"
/translation="MAAGDKAGGDDAAAAAPLLVSAAGRRRRCPGCLTERCKADA
GIPYLNFFYIWVCLCSFPRRRPAKKICDFSVKAMAHGHRR"
join(40347. .40358,40468. .40680)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87828.1"
/db_xref="GI:6498425"
```

CDS

```
/translation="MHPKNHLVAAATVILVLQSPLPSSSFSNRRRRRQCRPTIAIAT
ADGSTREVLPPPDRGGKGFRRMRMKEGGD"
complement(join(42025. .42083,43481. .43661))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87829.1"
/db_xref="GI:6498426"
/translation="MAAGDKAGGDDAAAAAPLLVPAAAGRRRRCPGCRTKERCEAHPG
IPYLNFFYIWIVCICASAVNRVEATEVMPTSTSVS"
join(47081. .47084,47194. .47313,47469. .47584)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87830.1"
/db_xref="GI:6498427"
/translation="MSRGRGRAPATACGSVRCYAAEVAPEQEDSARYALLLLRLRDSP
AATAPPKLARRIRDHDGVPLLERDEDLLSVGPTLD"
join(4812. .4822,49072. .49179,49283. .49438,49876. .50091,
50202. .50411,50488. .50625,50842. .50943,51105. .51170,
51274. .51358,51454. .51551)
/note="Similar to Arabidopsis thaliana chromosome II BAC
T27A16 sequence; hypothetical protein. (AC005496)"
/codon_start=1
/protein_id="BAA87831.1"
/db_xref="GI:6498428"
/translation="MLYLLPLSVSCRVPGSPPAPRSRRFLDPGGGRVGDGLGGVRVF
RRRALRGTDVRSNTSSSSSRKGRHDDARHDGGYGDDGDAGALLASVRRLLLSGSAQDD
AAEGEAEDEDEQGQFPKRWAIVFLCFSAFLLCMDRVNMSIALLPMSAEFGWNPQTVGL
IQSSFFWGYLITQIAGGIWADTVGGKTVLGFGVIWWSIATAITPFAAKLGLPFLLVTR
AFMGVGEGVAMPAMNNILSKWYSRSLALVYSGMYLGSVTGLAFSPLLIHNFGW
PSVFYSPGSLGVFWFSTWASKAYSSPLEDPGISAEEKLITSQTTGEPVKEIPWGLI
LSKPPVWALIVSHFCHNWGTFILLTWPTYNQVLKPNLTESGLFCVLPWLTMAVSAN
FGGWIADTLVSRGLSVTTVRKIMQSIGFLGPAFFLTQLSHIDSPAMAVLCMACSQGTD
AFSQSGLYSNHQDIGPRYAGVLLGLSNTAGVLAGVFGTAATGYILQHGSWDDVFKVSV
VLYLVGTLIVWNLFSTGEKIID"
complement(join(52616. .52685,53228. .53299,53395. .53466,
53657. .53800,54663. .54734,56026. .56234))
/note="ESTs C97644(C60871),AU081265(C60871) correspond to
a region of the predicted gene.; Similar to L.esculentum
LRP gene. (X95269)"
/codon_start=1
/protein_id="BAA87832.1"
/db_xref="GI:6498429"
/translation="MAAAWSPALAAVLLAAAVASASNEGDALYALRRALADPRGVLQ
SWDPTLVNPCTWHVTCDRAGRVTRLDLGNSNLSGHLAPELGHLEHLQYLELYKNNIQ
GTIPAELGSLKNLISLDLYNNNITGTIPKELGKLSSLVFLRLNDNSLNGPIPRDLAKI
SSLKVIDVSNNDLCGTIPTSGPFEHIPLNNFDKNPRLEGPELQGLATYDTNC"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87833.1"
/db_xref="GI:6498430"
/translation="MKLLKVDHPRSHTVFLYASLTSIALYTSLTIIVEKTGLAVQPE
KITGTSSLLTGSVPSKDRSRNRPGVNLADH"
complement(join(63689. .63842,63844. .64145))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA87834.1"
/db_xref="GI:6498431"
/translation="MRRSRARTDLRAAGRATAELRRRGRRSRGRGQEEYLGTASLGGG
EGSTSDEWEGRGGEGRGGAAAVRPAGEERRRHRAETKLGLSLVGRGAPHRQSGGHG
EGGRDEEGELMLVGGALRPEEEEEGHGRRRGRGEERRVCEREERER"
```

```
alignment_scores:
        Quality:     41.00                Length:       9
        Ratio:       4.556                Gaps:         0
        Percent Similarity: 88.889        Percent Identity: 100.000

alignment_block:
us-08-653-294-11 x AP000815/rev

Align seg 1/1  to reverse of: AP000815/rev   from:  1  to: 142418

   1 TyrArgLeuIleArgLeuAsnGlu 9
```

```
|||||||||||||||||||||||||||::
8117 TACCGCTTGTTAATTCGTTGAACCAA 8091

seq_name: gb_in1:CEC50B6

seq_documentation_block:
LOCUS       CEC50B6    41322 bp    DNA        INV     02-SEP-1999
DEFINITION  Caenorhabditis elegans cosmid C50B6, complete sequence.
ACCESSION   Z81050
VERSION     Z81050.1  GI:1627685
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41322)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
REFERENCE   2  (bases 1 to 41322)
AUTHORS     Percy,C.
TITLE       Direct Submission
JOURNAL     Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green; U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormaceclass=Sequence &object=C50B6
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C50B6.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C50B6 is at 1 in this sequence. The true
            right end of clone C50B6 is at 2293 in
            sequence 281524.
            The true left end of clone F32H5 is at 41219 in this sequence. The
            start of this sequence (1..101) overlaps with the end of sequence
            281040.
            The end of this sequence (41219..41322) overlaps with the start of
            sequence 281524.
FEATURES             Location/Qualifiers
     source          1..41322
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="v"
                     /clone="C50B6"
     gene            880..1777
                     /gene="C50B6.1"
     CDS             join(880..1077,1532..1639,1685..1777)
                     /gene="C50B6.1"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /protein_id="CAB02851.1"
                     /db_xref="GI:3875053"
                     /db_xref="SPTREMBL:O17688"
                     /translation="MSRHHVTKEGEKNGETLFTPEQLLRRGFAPLYLQEAYKNQEK
                     ERNWRKEIGGYKMKRQRLPAITGRQKAMLLQVMDRTKWPQGSVDRQKVADKVNLAYHQ
                     ITHQYRVHPENQMYMQNSKRQTREKITVSY"
     gene            complement(2944..4842)
                     /gene="C50B6.2"
     CDS             complement(join(2944..3651,3698..4050,4106..4395,
                     4445..4842)
                     /gene="C50B6.2"
                     /note="predicted using Genefinder; Similarity to Xenopus
                     histone binding protein (SW:P06180); cDNA EST EMBL:Z14755
                     comes from this gene; cDNA EST EMBL:Z14820 comes from this
                     gene; cDNA EST EMBL:T01234 comes from this gene; cDNA EST
                     EMBL:T02164 comes from this gene; cDNA EST EMBL:M79494
                     comes from this gene; cDNA EST CEMSA18R comes from this
                     gene; cDNA EST EMBL:D33193 comes from this gene; cDNA EST
                     EMBL:D35959 comes from this gene; cDNA EST EMBL:D71778
                     comes from this gene; cDNA EST EMBL:D72307 comes from this
                     gene; cDNA EST EMBL:D72719 comes from this gene; cDNA EST
                     EMBL:D74467 comes from this gene; cDNA EST EMBL:D75157
                     comes from this gene; cDNA EST EMBL:D75631 comes from this
                     gene; cDNA EST EMBL:D64450 comes from this gene; cDNA EST
                     EMBL:D64882 comes from this gene; cDNA EST EMBL:D65720
                     comes from this gene; cDNA EST EMBL:D65951 comes from this
                     gene; cDNA EST EMBL:D67440 comes from this gene; cDNA EST
                     EMBL:D68140 comes from this gene; cDNA EST EMBL:D69258
                     comes from this gene; cDNA EST EMBL:D69570 comes from this
                     gene; cDNA EST EMBL:C12197 comes from this gene; cDNA EST
                     EMBL:C10518 comes from this gene: cDNA EST yk670d2.3 comes
                     from this gene"
                     /codon_start=1
                     /protein_id="CAB02850.1"
                     /db_xref="GI:3875052"
                     /db_xref="SPTREMBL:O17687"
                     /translation="MSAFADSTNIIAHDKETIAKNVDSTGEEKVAMTNVAEETTPSTD
                     SATTNEKTASENSPTEAAASDVKDVAEASTGAQETNTSSETPAVVEDEKASMEVSIGT
                     SIEDSETSQKEIEGELVVPENPVEEISPEQKRKSLSENIIAGRRFMRTDVFDKAVDVL
                     SVAASIAAEVFGDAHEDTFEANFLYGKALLVGKLEDRVLANALTDVPKMAEGEEEVQ
                     DGIVENPEDVPQDERAEIKQKVEEALGVASEEPETVADEAVKTEQKEAEEDSEVKDVE
                     NSDEQNQEEEVVENEATVDPTEDVEMEGVEEVNDEKEVGGTAEVDGEAEEDDEAEDD
                     EADSMKLAWELLETSRCIADKKAASLAAESTVDGEAIKMWKLNLADVLTISLGEHGIAD
                     SKYEQAQKDLTEAISIQTVHLPATSRVIANTHLLAKAFSSDSLFEQAAAHFNDTKNI
                     LIAKSDELKQLETVSDDAKSDIQSEIKELDGIIPELDAFIVDARASAEQTEKLKSDA
                     KKELEGAAAMLAKVADVTEATDISSMVRRPTKRPASEEAPETKKRKSGDGVDVAVVT
                     EEQVVSTEEPTPINE"
     gene            6195..8084
                     /gene="C50B6.3"
     CDS             join(6195..6268,6321..6483,6532..7116,7207..7428,
                     7495..7971,8025..8084)
                     /gene="C50B6.3"
                     /note="predicted using Genefinder"
                     /codon_start=1
                     /protein_id="CAB02852.1"
                     /db_xref="GI:3875054"
                     /db_xref="SPTREMBL:O17689"
                     /translation="MVEGDVDESASGTSGTNKKLLFTKKPSVWKDFDNWINDEPENRY
                     DLFQVVKSAMLQSGYTTILMDQVTNGADELRISLEYSNFIKIVNSTKLVVGKEQCP
                     PSNVFTLLAEIFANTPGNTSEVGRISTWLTSHLGALLHNDVIWKIHFFDPDLFRSVYW
                     QLIFTIKLAPGDTENLEEDENYAKLLFSCFITAVMVALWDHEMSFNSICPDYLRPET
                     ASEYMVMLISSPPFRSLSQFLFGLHLLGKYQSEGGCVVVREEAYIAEIRQNDEEKRQ
                     RPRKYPASAIKKRKKKTPRSSPKKMKSKESPINHQKEPIDEQKPSTLPIYSVATLKP
                     RRKVVKTADEVGLCAPIFVMQSELLKKFREEVQRRYAEGSSASDQPRVRNMYEAYDN
                     IYHINRLSANEGPRILTSDQKLVMQQYKTTFRQGPTFAEETESDVEEEEKKVVEVVT
                     AKVIKGSAKSSKKFKRRY"
     gene            complement(9672..10594)
                     /gene="C50B6.4"
     CDS             complement(join(9672..10406,10454..10594))
```

```
38600 TATAGGATACTGCTCCGATTGAATGAA 38626

seq_name: gb_htg7:AC017780

seq_documentation_block:
LOCUS       AC017780    60818 bp    DNA              HTG      09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017780
VERSION     AC017780.1  GI:6553410
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 60818)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212069 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..60818
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    15775 a  13372 c  13641 g  18030 t
ORIGIN

alignment_scores:
     Quality:      40.00              Length:          9
     Ratio:         4.444             Gaps:            0
Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
US-08-653-294-11 x AC017780

Align seg 1/1  to: AC017780   from: 1   to: 60818

    1 TyrArgLeuIleArgLeuAsnGlu 9
      ::|||||||||||||||||||||||
24132 TCCCGTTACTTATTCGGTTGAACGAG 24158

seq_name: gb_htg3:AC009355

seq_documentation_block:
LOCUS       AC009355    66958 bp    DNA              HTG      20-SEP-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR28N20 (D919) RPCI-98
            28.N.20 map 25C-25C strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***,76 unordered pieces.
ACCESSION   AC009355
VERSION     AC009355.3  GI:5512617
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 66958)
AUTHORS     Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R.J., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hunmasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
```

```
                     /gene="C50B6.4"
                     /note="predicted using Genefinder; similar to collagen;
                     cDNA EST EMBL:D65564 comes from this gene; cDNA EST
                     EMBL:D69046 comes from this gene; cDNA EST yk366b12.3
                     comes from this gene; cDNA EST yk366b12.5 comes from this
                     gene"
                     /codon_start=1
                     /protein_id="CAB02849.1"
                     /db_xref="GI:3875051"
                     /db_xref="SPTREMBL:O17686"
                     /translation="MSAKTLVYGASVLSGFAILGCVFTVGYIFNDINEFYEQTMETMD
                     EFKLNERGAWHGMVERTRAPSEILFGRAKRQAGQCNCGAQSSGCPAGPPGPPGAPGAP
                     GDDGHAGEAGKFGTAGVAVGIVSEGGFCIKCPAGEPGPAGEAGAPGEAGAPGDPGPGQDG
                     QGGQPGPAGPAGDGAGAPGAGAPGDGPGAPGQDGDGQRSTGTPGAAGAPGAPGAPGAGD
                     GQPGSDGPGEAGAPGAPGPGPGQPGQDGEAGAPGQDGQPGSDAAYCPCPARSVAV
                     QRSVSRRRASKVVVA"
gene            11061..12677
                     /gene="C50B6.5"
CDS             join(11061,11221,11307..11452,11897..12333,12385..12463,
                     12511..12677)
                     /gene="C50B6.5"
                     /codon_start=1
                     /protein_id="CAB02854.1"
                     /db_xref="GI:3875056"
                     /db_xref="SPTREMBL:P91980"
                     /translation="MDCSLISTAEYLDVLHFLFIVSFPIYTVAIIALFRTKSTYFETY
                     KHFLVWHTASNLISEIYNAWFLAPKVHLPYPLIRFTAIMTOLGFSGLFQFYTINALIH
                     QTGYSIIEMYMRFKASTYNFQSTCFVYLQINLYIIRITLVLFFVVNITTYNISLGQ
                     QIISKQNLLIQHPEAPWLVNCDSVVVAAPFTDPISMFNVVWIVIIFVASTSFSTTI
                     YLQKHLSKSEHSPAVLRMHRMLLITLFVQTAIHAVMLGIPNSMFIYAVFFEARHEFL
                     AKIAFCCLITYHGLASTIAMMTLTKPIKITILQMLRCHALKNSVSTVQYSSQNT"
gene            13581..14952
                     /gene="C50B6.6"
CDS             join(13581,13654,13709..13834,13884..14029,14205..14272,
                     14327..14464,14708..14952)
                     /gene="C50B6.6"
                     /note="similar to RNA binding protein"
                     /codon_start=1
                     /protein_id="CAB02855.1"
                     /db_xref="GI:3875057"
                     /db_xref="SPTREMBL:P91981"
                     /translation="MGTTAPVVITPSSEGCLSPAPDPYKSMMHFTHVLTIPLYLTAIY
                     ALINNCPKTLKEYRKYLLWHTIGNLLFELYISLFMLPVTYLPYPVFRGAGFLKYLDIS
                     GLIQFYLVLVICLJLFCFATLPRCHPKQHYKQDLYDKNPGISIHVLCNTVVTAPPLID
                     PVFTPLMTLIMAMLLAATIIPQTFLRTIWKKLDQLSKHLSKKTIQLQKMLITSLFIQA
                     VIHGVMLGAPLIGFIYAVFVFLPYNYIAYMLLLISFHGSFSTIAMIAFTKPIREGVK
                     SLFQSLFFLILSEPRTKIMVVTSTSASTKQRDSIFMSNKNSFLSV"
complement(15021..17567)
                     /gene="C50B6.7"
CDS             complement(join(15021..15502,15568..15753,15799..15902,
                     15954..16171,16221..16634,16680..16871,16918..17122,
                     17227..17567))
                     /gene="C50B6.7"
                     /note="Similarity to Shrimp amylase (TR:Q26193); cDNA EST
                     EMBL:Z14343 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB02856.1"
                     /db_xref="GI:3875058"
                     /db_xref="SPTREMBL:P91982"
                     /translation="MLKHLFLFIKSSIAYNFYWYDKTQTLQNRQTMVHLFEWKWAD

alignment_scores:
     Quality:      40.00              Length:          9
     Ratio:         4.444             Gaps:            0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x CEC50B6

Align seg 1/1  to: CEC50B6   from: 1   to: 41322

    1 TyrArgLeuIleArgLeuAsnGlu 9
      |||||::::|||:::||||||||||
```

```
            Rubin,G.M.
TITLE     Sequencing of Drosophila melanogaster
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 66958)
AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
          Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
          Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
          Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
          Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
          Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
          Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
          Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
          Rubin,G.M.
TITLE     Direct Submission
JOURNAL   Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
          Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT   On Sep 20, 1999 this sequence version replaced gi:5748857.
          For further information about this sequence, including its location
          and relationship to other sequences, please visit our sequence
          archive Web site (http://www.fruitfly.org/sequence/) or send email
          to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
          the following cutoffs: length >= 200 bases.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 76 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *     1  1075: contig of 1075 bp in length
        *  1076  1155: gap of unknown length
        *  1156  1835: contig of 680 bp in length
        *  1836  1915: gap of unknown length
        *  1916  2905: contig of 990 bp in length
        *  2906  2985: gap of unknown length
        *  2986  3617: contig of 632 bp in length
        *  3618  3697: gap of unknown length
        *  3698  4682: contig of 985 bp in length
        *  4683  4762: gap of unknown length
        *  4763  5235: contig of 473 bp in length
        *  5236  5315: gap of unknown length
        *  5316  5956: contig of 641 bp in length
        *  5957  6036: gap of unknown length
        *  6037  6655: contig of 619 bp in length
        *  6656  6735: gap of unknown length
        *  6736  7335: contig of 600 bp in length
        *  7336  7415: gap of unknown length
        *  7416  8119: contig of 704 bp in length
        *  8120  8199: gap of unknown length
        *  8200  8769: contig of 570 bp in length
        *  8770  8849: gap of unknown length
        *  8850  9419: contig of 570 bp in length
        *  9420  9499: gap of unknown length
        *  9500  9915: contig of 416 bp in length
        *  9916  9995: gap of unknown length
        *  9996 11078: contig of 1083 bp in length
        * 11079 11158: gap of unknown length
        * 11159 12139: contig of 981 bp in length
        * 12140 12219: gap of unknown length
        * 12220 12951: contig of 732 bp in length
        * 12952 13031: gap of unknown length
        * 13032 13776: contig of 745 bp in length
        * 13777 13856: gap of unknown length
        * 13857 15280: contig of 1424 bp in length
        * 15281 15360: gap of unknown length
        * 15361 16485: contig of 1125 bp in length
        * 16486 16565: gap of unknown length
        * 16566 17573: contig of 1008 bp in length
        * 17574 17653: gap of unknown length
        * 17654 18313: contig of 660 bp in length
        * 18314 18393: gap of unknown length
        * 18394 19168: contig of 775 bp in length

        * 19169 19248: gap of unknown length
        * 19249 20162: contig of 914 bp in length
        * 20163 20242: gap of unknown length
        * 20243 21204: contig of 962 bp in length
        * 21205 21284: gap of unknown length
        * 21285 22698: contig of 1414 bp in length
        * 22699 22778: gap of unknown length
        * 22779 24100: contig of 1322 bp in length
        * 24101 24180: gap of unknown length
        * 24181 25936: contig of 1756 bp in length
        * 25937 26016: gap of unknown length
        * 26017 27558: contig of 1542 bp in length
        * 27559 27638: gap of unknown length
        * 27639 29229: contig of 1591 bp in length
        * 29230 29309: gap of unknown length
        * 29310 30695: contig of 1386 bp in length
        * 30696 30775: gap of unknown length
        * 30776 32559: contig of 1784 bp in length
        * 32560 32639: gap of unknown length
        * 32640 33600: contig of 961 bp in length
        * 33601 33680: gap of unknown length
        * 33681 35311: contig of 1631 bp in length
        * 35312 35391: gap of unknown length
        * 35392 36278: contig of 887 bp in length
        * 36279 36358: gap of unknown length
        * 36359 38465: contig of 2107 bp in length
        * 38466 38545: gap of unknown length
        * 38546 40513: contig of 1968 bp in length
        * 40514 40593: gap of unknown length
        * 40594 41148: contig of 555 bp in length
        * 41149 41228: gap of unknown length
        * 41229 41878: contig of 650 bp in length
        * 41879 41958: gap of unknown length
        * 41959 42286: contig of 328 bp in length
        * 42287 42366: gap of unknown length
        * 42367 42957: contig of 591 bp in length
        * 42958 43037: gap of unknown length
        * 43038 43592: contig of 555 bp in length
        * 43593 43672: gap of unknown length
        * 43673 44269: contig of 597 bp in length
        * 44270 44349: gap of unknown length
        * 44350 45099: contig of 750 bp in length
        * 45100 45179: gap of unknown length
        * 45180 45825: contig of 646 bp in length
        * 45826 45905: gap of unknown length
        * 45906 46485: contig of 580 bp in length
        * 46486 46565: gap of unknown length
        * 46566 47171: contig of 606 bp in length
        * 47172 47251: gap of unknown length
        * 47252 47832: contig of 581 bp in length
        * 47833 47912: gap of unknown length
        * 47913 48534: contig of 622 bp in length
        * 48535 48614: gap of unknown length
        * 48615 49185: contig of 571 bp in length
        * 49186 49265: gap of unknown length
        * 49266 49805: contig of 540 bp in length
        * 49806 49885: gap of unknown length
        * 49886 50509: contig of 624 bp in length
        * 50510 50589: gap of unknown length
        * 50590 51193: contig of 604 bp in length
        * 51194 51273: gap of unknown length
        * 51274 51887: contig of 614 bp in length
        * 51888 51967: gap of unknown length
        * 51968 52590: contig of 623 bp in length
        * 52591 52670: gap of unknown length
        * 52671 53058: contig of 388 bp in length
        * 53059 53138: gap of unknown length
        * 53139 53656: contig of 518 bp in length
        * 53657 53736: gap of unknown length
        * 53737 54452: contig of 716 bp in length
        * 54453 54532: gap of unknown length
        * 54533 55105: contig of 573 bp in length
        * 55106 55185: gap of unknown length
        * 55186 ...
```

```
*   55186   55736:  contig of 551 bp in length
*   55737   55816:  gap of unknown length
*   55817   56383:  contig of 567 bp in length
*   56384   56463:  gap of unknown length
*   56464   57073:  contig of 610 bp in length
*   57074   57153:  gap of unknown length
*   57154   57718:  contig of 565 bp in length
*   57719   57798:  gap of unknown length
*   57799   58374:  contig of 576 bp in length
*   58375   58454:  gap of unknown length
*   58455   58951:  contig of 497 bp in length
*   58952   59031:  gap of unknown length
*   59032   59665:  contig of 634 bp in length
*   59666   59745:  gap of unknown length
*   59746   60315:  contig of 570 bp in length
*   60316   60395:  gap of unknown length
*   60396   60964:  contig of 569 bp in length
*   60965   61044:  gap of unknown length
*   61045   61568:  contig of 524 bp in length
*   61569   61648:  gap of unknown length
*   61649   62198:  contig of 550 bp in length
*   62199   62278:  gap of unknown length
*   62279   62868:  contig of 590 bp in length
*   62869   62948:  gap of unknown length
*   62949   63560:  contig of 612 bp in length
*   63561   63640:  gap of unknown length
*   63641   64213:  contig of 573 bp in length
*   64214   64293:  gap of unknown length
*   64294   64860:  contig of 567 bp in length
*   64861   64940:  gap of unknown length
*   64941   65582:  contig of 642 bp in length
*   65583   65662:  gap of unknown length
*   65663   66213:  contig of 551 bp in length
```

```
alignment_scores:
       Quality:    40.00        Length:       9
       Ratio:     4.444         Gaps:         0
Percent Similarity: 100.000     Percent Identity: 88.889

alignment_block:
US-08-653-294-11 x AC009355/rev   ..

Align seg 1/1  to reverse of: AC009355  from: 1  to: 66958

   1 TyrArgLeuLeuIleArgLeuAsnGlu 9
     :::||||||||||||||||||||||
46014 TTCCGGTTACTTATTCGGTTGAACGAG 45988

seq_name: gb_pr3:AC005365

seq_documentation_block:
LOCUS       AC005365    86130 bp   DNA             PRI    01-AUG-1998
DEFINITION  Homo sapiens chromosome 16, P1 clone 79-2A (LANL), complete
            sequence.
ACCESSION   AC005365
VERSION     AC005365.1  GI:3367509
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86130)
AUTHORS     Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatun,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
TITLE       Sequencing of Human Chromosome 16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 86130)
AUTHORS     Ricke,D.O.
TITLE       Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
```

```
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 86130)
AUTHORS     Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatun,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
TITLE       Direct Submission
JOURNAL     Submitted (01-AUG-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1..86130
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="79-2A"
     misc_feature    2000..2142
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   2185..2497
                     /rpt_family="Alu"
     repeat_region   7853..8448
                     /rpt_family="Alu"
     repeat_region   9717..10001
                     /rpt_family="Alu"
     repeat_region   10542..10563
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(10546..10672)
                     /rpt_family="Alu"
     misc_feature    11817..12730
     misc_feature    complement(14804..14902)
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    15869..16008
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   16275..16548
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   17124..17355
                     /rpt_family="Alu"
     repeat_region   complement(18448..18747)
                     /rpt_family="MER33"
     repeat_region   19083..19373
                     /rpt_family="Alu"
     repeat_region   19357..19380
                     /note="(A)24"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   20307..20329
                     /note="(T)23"
                     /rpt_type=tandem
                     /rpt_unit=T
     misc_feature    20836..21048
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   21252..21358
                     /rpt_family="Alu"
     repeat_region   complement(22142..22428)
                     /rpt_family="Alu"
     repeat_region   27568..27837
                     /rpt_family="Alu"
     misc_feature    28059..28174
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   28975..29221
                     /rpt_family="Alu"
     repeat_region   29408..29459
                     /rpt_family="MIR"
     repeat_region   30146..30276
                     /rpt_family="Alu"
     misc_feature    30513..30631
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   30919..31186
                     /rpt_family="Alu"
     repeat_region   31608..32152
```

```
repeat_region     /rpt_family="Alu"
                  complement(32560..32851)
repeat_region     /rpt_family="Alu"
                  33192..33293
repeat_region     /rpt_family="Alu"
                  35811..35929
misc_feature      /note="GRAIL 2 excellent exon, frame 2"
repeat_region     36035..36312
                  /rpt_family="Alu"
repeat_region     complement(36626..36748)
                  /rpt_family="L1"
repeat_region     complement(37173..37468)
                  /rpt_family="Alu"
repeat_region     38519..38818
                  /rpt_family="Alu"
repeat_region     40555..40673
                  /rpt_family="MIR"
repeat_region     41506..41532
                  /note="(A)27"
                  /rpt_type=tandem
                  /rpt_unit=A
repeat_region     45596..45892
                  /rpt_family="Alu"
repeat_region     46081..46366
                  /rpt_family="Alu"
misc_feature      46477..46721
                  /note="GRAIL 2 excellent exon, frame 0"
repeat_region     47484..47760
                  /rpt_family="L1"
misc_feature      50106..50312
                  /note="GRAIL 2 excellent exon, frame 2"
repeat_region     50512..50788
                  /rpt_family="Alu"
repeat_region     50789..51068
                  /rpt_family="Alu"
repeat_region     complement(52275..52538)
                  /rpt_family="Alu"
repeat_region     complement(53058..53720)
                  /rpt_family="L1"
repeat_region     54695..54976
                  /rpt_family="Alu"
repeat_region     55556..55829
                  /rpt_family="Alu"
repeat_region     56565..56848
                  /rpt_family="Alu"
misc_feature      complement(57163..57314)
                  /note="GRAIL 2 excellent exon, frame 1"
repeat_region     58926..59183
                  /rpt_family="Alu"
repeat_region     59490..60014
                  /rpt_family="Alu"
repeat_region     complement(60362..60655)
                  /rpt_family="Alu"
repeat_region     60988..61273
                  /rpt_family="Alu"
repeat_region     61417..61757
                  /rpt_family="Alu"
repeat_region     complement(62028..62141)
                  /rpt_family="Alu"
repeat_region     complement(63160..63454)
                  /rpt_family="Alu"
repeat_region     complement(63604..63875)
                  /rpt_family="Alu"
repeat_region     complement(63894..64098)
                  /rpt_family="MER1"
misc_feature      64463..64588
                  /note="GRAIL 2 excellent exon, frame 2"
repeat_region     64707..64992
                  /rpt_family="Alu"
misc_feature      66355..66492
                  /note="GRAIL 2 excellent exon, frame 0"
repeat_region     complement(66679..66796)
                  /rpt_family="Alu"

repeat_region     66995..67262
                  /rpt_family="Alu"
repeat_region     67741..68036
                  /rpt_family="Alu"
repeat_region     complement(69213..69456)
                  /rpt_family="Alu"
repeat_region     70077..70906
                  /rpt_family="LTR1"
repeat_region     71602..72045
                  /rpt_family="L1"
repeat_region     72758..72876
                  /rpt_family="L1"
repeat_region     complement(72760..72985)
                  /rpt_family="MER46"
repeat_region     complement(73181..73243)
                  /rpt_family="MER46"
repeat_region     complement(74246..74532)
                  /rpt_family="MIR"
repeat_region     74601..74899
                  /rpt_family="Alu"
repeat_region     74915..75035
                  /rpt_family="L1"
repeat_region     75843..76122
                  /rpt_family="Alu"
repeat_region     76105..76145
                  /note="(A)41"
                  /rpt_type=tandem
                  /rpt_unit=A
repeat_region     76163..76652
                  /rpt_family="Alu"
repeat_region     76635..76655
                  /note="(A)21"
```

```
alignment_scores:
              Quality:   40.00           Length:       10
                Ratio:    4.444           Gaps:          0
  Percent Similarity:   90.000   Percent Identity:   80.000

alignment_block:
  US-08-653-294-11 x AC005365

Align seg 1/1  to: AC005365    from: 1  to: 86130

      1 TyrArgLeuIleArgLeuAsnGluArg 10
        |||||||||::||||||||||||  |||
  23775 TACAGACTTGTCATCAGATTGAATCCCAGA 23804

seq_name: gb_htg1:HS1141E20

seq_documentation_block:
LOCUS       HS1141E20       97906 bp    DNA              HTG       23-NOV-1999
DEFINITION  Homo sapiens chromosome 6 clone RP5-1141E20, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL109912
VERSION     AL109912.4  GI:5870369
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97906)
AUTHORS     Sims,S.
TITLE       Direct Submission
JOURNAL     Submitted (10-SEP-1999) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 12, 1999 this sequence version replaced gi:5777438.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
```

COMMENT

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.berkeley.edu/) or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 81 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1    594:  contig of 594 bp in length
*      595    674:  gap of unknown length
*      675   1814:  contig of 1140 bp in length
*     1815   1894:  gap of unknown length
*     1895   2545:  contig of 651 bp in length
*     2546   2625:  gap of unknown length
*     2626   3091:  contig of 466 bp in length
*     3092   3171:  gap of unknown length
*     3172   4398:  contig of 1227 bp in length
*     4399   4478:  gap of unknown length
*     4479   5421:  contig of 943 bp in length
*     5422   5501:  gap of unknown length
*     5502   6196:  contig of 695 bp in length
*     6197   6276:  gap of unknown length
*     6277   7008:  contig of 732 bp in length
*     7009   7088:  gap of unknown length
*     7089   7770:  contig of 682 bp in length
*     7771   7850:  gap of unknown length
*     7851   8573:  contig of 723 bp in length
*     8574   8653:  gap of unknown length
*     8654   9411:  contig of 758 bp in length
*     9412   9491:  gap of unknown length
*     9492  10299:  contig of 808 bp in length
*    10300  10379:  gap of unknown length
*    10380  11647:  contig of 1268 bp in length
*    11648  11727:  gap of unknown length
*    11728  12578:  contig of 851 bp in length
*    12579  12658:  gap of unknown length
*    12659  13474:  contig of 816 bp in length
*    13475  13554:  gap of unknown length
*    13555  14454:  contig of 900 bp in length
*    14455  14534:  gap of unknown length
*    14535  15301:  contig of 767 bp in length
*    15302  15381:  gap of unknown length
*    15382  16820:  contig of 1439 bp in length
*    16821  16900:  gap of unknown length
*    16901  17817:  contig of 917 bp in length
*    17818  17897:  gap of unknown length
*    17898  19666:  contig of 1769 bp in length
*    19667  19746:  gap of unknown length
*    19747  21069:  contig of 1323 bp in length
*    21070  21149:  gap of unknown length
*    21150  22084:  contig of 935 bp in length
*    22085  22164:  gap of unknown length
*    22165  23539:  contig of 1375 bp in length
*    23540  23619:  gap of unknown length
*    23620  25170:  contig of 1551 bp in length
*    25171  25250:  gap of unknown length
*    25251  26158:  contig of 908 bp in length
*    26159  26238:  gap of unknown length
*    26239  27461:  contig of 1223 bp in length
*    27462  27541:  gap of unknown length
*    27542  28048:  contig of 507 bp in length
*    28049  28128:  gap of unknown length
*    28129  29477:  contig of 1349 bp in length
*    29478  29557:  gap of unknown length
*    29558  31107:  contig of 1550 bp in length
*    31108  31187:  gap of unknown length
*    31188  31856:  contig of 669 bp in length
*    31857  31936:  gap of unknown length
```

segments, Unfinished: dJ1141E20 Contig_ID: 00340 acc=AL109912
Length: 64235 bp Unfinished: dJ1141E20 Contig_ID: 00618
acc=AL109912 Length: 18869 bp Unfinished: dJ1141E20 Contig_ID:
00679 acc=AL109912 Length: 13202 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES   Location/Qualifiers
source   1..97906
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="6"
                  /clone_lib="RPCI-5"
                  /clone="RP5-1141E20"

BASE COUNT   30195 a   18298 c   18016 g   29795 t   1602 others
ORIGIN

alignment_scores:
Quality: 40.00   Length: 10
Ratio: 4.444   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 80.000

alignment_block:
US-08-653-294-11 x HS1141E20/rev

Align seg 1/1   to reverse of: HS1141E20   from: 1 to: 97906

```
    1 TyrArgLeuLeuArgLeuAsnGluArg 10
      ||||||||:::||||||||| |||
48886 TACAGACTGTCATCAGATTGAATCCCAGA 48857
```

seq_name: gb_htg3:AC008324

seq_documentation_block:
LOCUS   AC008324   122061 bp   DNA   HTG   06-AUG-1999
DEFINITION Drosophila melanogaster chromosome 2 clone BACR25K01 (D854) RPCI-98
            25.K.1 map 25C-25D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 81 unordered pieces.
ACCESSION   AC008324
VERSION   AC008324.1 GI:5670415
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 122061)
AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE   Sequencing of Drosophila melanogaster
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 122061)
AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA

```
*  31937  33142: contig of 1206 bp in length
*  33143  33222: gap of unknown length
*  33223  34713: contig of 1491 bp in length
*  34714  34793: gap of unknown length
*  34794  36055: contig of 1262 bp in length
*  36056  36135: gap of unknown length
*  36136  37830: contig of 1695 bp in length
*  37831  37910: gap of unknown length
*  37911  39270: contig of 1360 bp in length
*  39271  39350: gap of unknown length
*  39351  40199: contig of 849 bp in length
*  40200  40279: gap of unknown length
*  40280  42203: contig of 1924 bp in length
*  42204  42283: gap of unknown length
*  42284  44017: contig of 1734 bp in length
*  44018  44097: gap of unknown length
*  44098  45611: contig of 1514 bp in length
*  45612  45691: gap of unknown length
*  45692  47034: contig of 1343 bp in length
*  47035  47114: gap of unknown length
*  47115  48858: contig of 1744 bp in length
*  48859  48938: gap of unknown length
*  48939  50521: contig of 1583 bp in length
*  50522  50601: gap of unknown length
*  50602  52996: contig of 2395 bp in length
*  52997  53076: gap of unknown length
*  53077  54627: contig of 1551 bp in length
*  54628  54707: gap of unknown length
*  54708  56095: contig of 1388 bp in length
*  56096  56175: gap of unknown length
*  56176  57722: contig of 1547 bp in length
*  57723  57802: gap of unknown length
*  57803  60670: contig of 2868 bp in length
*  60671  60750: gap of unknown length
*  60751  62436: contig of 1686 bp in length
*  62437  62516: gap of unknown length
*  62517  64900: contig of 2384 bp in length
*  64901  64980: gap of unknown length
*  64981  67157: contig of 2177 bp in length
*  67158  67237: gap of unknown length
*  67238  70347: contig of 3110 bp in length
*  70348  70427: gap of unknown length
*  70428  73471: contig of 3044 bp in length
*  73472  73551: gap of unknown length
*  73552  76116: contig of 2565 bp in length
*  76117  76196: gap of unknown length
*  76197  79584: contig of 3388 bp in length
*  79585  79664: gap of unknown length
*  79665  82547: contig of 2883 bp in length
*  82548  82627: gap of unknown length
*  82628  85279: contig of 2652 bp in length
*  85280  85359: gap of unknown length
*  85360  88875: contig of 3516 bp in length
*  88876  88955: gap of unknown length
*  88956  93882: contig of 4927 bp in length
*  93883  93962: gap of unknown length
*  93963  98575: contig of 4613 bp in length
*  98576  98655: gap of unknown length
*  98656 105323: contig of 6668 bp in length
* 105324 105403: gap of unknown length
* 105404 106093: contig of 690 bp in length
* 106094 106173: gap of unknown length
* 106174 106901: contig of 728 bp in length
* 106902 106981: gap of unknown length
* 106982 107613: contig of 631 bp in length
* 107613 107692: gap of unknown length
* 107693 108426: contig of 734 bp in length
* 108427 108506: gap of unknown length
* 108507 109243: contig of 737 bp in length
* 109244 109323: gap of unknown length
* 109324 110081: contig of 758 bp in length
* 110082 110161: gap of unknown length
* 110162 110890: contig of 729 bp in length
* 110891 110970: gap of unknown length
* 110971 111686: contig of 716 bp in length
* 111687 111766: gap of unknown length
* 111767 112474: contig of 708 bp in length
* 112475 112554: gap of unknown length
* 112555 113312: contig of 758 bp in length
* 113313 113392: gap of unknown length
* 113393 114092: contig of 700 bp in length
* 114093 114172: gap of unknown length
* 114173 114930: contig of 758 bp in length
* 114931 115010: gap of unknown length
* 115011 115817: contig of 807 bp in length
* 115818 115897: gap of unknown length
* 115898 116596: contig of 699 bp in length
* 116597 116676: gap of unknown length
* 116677 117456: contig of 780 bp in length
* 117457 117536: gap of unknown length
```

```
alignment_scores:
       Quality:  40.00            Length:    9
       Ratio:     4.444           Gaps:      0
Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
US-08-653-294-11 x AC008324/rev

Align seg 1/1 to reverse of: AC008324   from: 1  to: 122061

    1 TyrArgLeuIleArgLeuAsnGlu 9
      :::||||||||||||||||||||||
23076 TTCCGGTTACTTATTCGGTTGAACGAG 23050

seq_name: gb_pr3:HSJ1112D6

seq_documentation_block:
LOCUS       HSJ1112D6   135305 bp    DNA        PRI       23-NOV-1999
DEFINITION  Human DNA sequence from clone 1112D6 on chromosome 6q21-22.2,
            complete sequence.
ACCESSION   AL080317
VERSION     AL080317.11  GI:5830430
KEYWORDS    HTG; CpG Island.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 135305)
AUTHORS     Patel,R.
TITLE       Direct Submission
JOURNAL     Submitted (09-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 1999 this sequence version replaced gi:5791529.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is
            the entire insert of clone 1112D6. This sequence has been finished
            according to sequence map criteria as follows. An attempt is made
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
```

```
          1112D6 is from the library RPCI-5 constructed at the Roswell Park
          Cancer Institute by the group of Pieter de Jong. For further
          details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES       Location/Qualifiers
     source    1..135305
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="6"
               /map="q21-22.2"
               /clone_lib="RPCI-5"
               /clone="RP5-1112D6"
BASE COUNT    41171 a   26002 c   26091 g   42041 t
ORIGIN

alignment_scores:
     Quality:         40.00          Length:       10
     Ratio:           4.444          Gaps:          0
     Percent Similarity: 90.000      Percent Identity: 80.000

alignment_block:
US-08-653-294-11 x HSJ1112D6/rev

Align seg 1/1  to reverse of: HSJ1112D6  from: 1  to: 135305

     1 TyrArgLeuIleArgLeuAsnGluArg 10
       |||||||||::|||||||||||  |||
 99309 TACAGACTTGTCATCAGATTGAATCCCAGA 99280

seq_name: gb_htg3:AC008327

seq_documentation_block:
LOCUS       AC008327   185469 bp    DNA          HTG       20-AUG-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR14K04 (D859) RPCI-98
            14.K.4 map 27C-27C strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 188 unordered pieces.
ACCESSION   AC008327
VERSION     AC008327.2  GI:5748862
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 185469)
AUTHORS     Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blaze],R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE       Sequencing of Drosophila melanogaster
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 185469)
AUTHORS     Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blaze],R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 20, 1999 this sequence version replaced gi:5670411.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
```

```
          to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
          the following cutoffs: length >= 200 bases.
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 188 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*      672     673: contig of 672 bp in length
*      752     753: gap of unknown length
*     1942    1943: contig of 1190 bp in length
*     2022    2023: gap of unknown length
*     2902    2903: contig of 880 bp in length
*     2982    2983: gap of unknown length
*     4320    4321: contig of 1338 bp in length
*     4400    4401: gap of unknown length
*     5184    5185: contig of 784 bp in length
*     5264    5265: gap of unknown length
*     6058    6059: contig of 794 bp in length
*     6138    6139: gap of unknown length
*     6857    6858: contig of 719 bp in length
*     6937    6938: gap of unknown length
*     8116    8117: contig of 1179 bp in length
*     8196    8197: gap of unknown length
*     8893    8894: contig of 697 bp in length
*     8973    8974: gap of unknown length
*     9623    9624: contig of 650 bp in length
*     9703    9704: gap of unknown length
*    10744   10745: contig of 1041 bp in length
*    10824   10825: gap of unknown length
*    11514   11515: contig of 690 bp in length
*    11594   11595: gap of unknown length
*    12380   12381: contig of 786 bp in length
*    12460   12461: gap of unknown length
*    13282   13283: contig of 822 bp in length
*    13362   13363: gap of unknown length
*    13938   13939: contig of 576 bp in length
*    14018   14019: gap of unknown length
*    15240   15241: contig of 1222 bp in length
*    15320   15321: gap of unknown length
*    16141   16142: contig of 821 bp in length
*    16221   16222: gap of unknown length
*    16940   16941: contig of 719 bp in length
*    17020   17021: gap of unknown length
*    17930   17931: contig of 910 bp in length
*    18010   18011: gap of unknown length
*    18792   18793: contig of 782 bp in length
*    18872   18873: gap of unknown length
*    19592   19593: contig of 720 bp in length
*    19672   19673: gap of unknown length
*    20826   20827: contig of 1154 bp in length
*    20906   20907: gap of unknown length
*    21710   21711: contig of 804 bp in length
*    21790   21791: gap of unknown length
*    22994   22995: contig of 1204 bp in length
*    23074   23075: gap of unknown length
*    24155   24156: contig of 1081 bp in length
*    24235   24236: gap of unknown length
*    25381   25382: contig of 1146 bp in length
*    25461   25462: gap of unknown length
*    26295   26296: contig of 834 bp in length
*    26375   26376: gap of unknown length
*    27786   27787: contig of 1411 bp in length
*    27866   27867: gap of unknown length
*    28693   28694: contig of 827 bp in length
*    28773   28774: gap of unknown length
*    29479   29480: contig of 706 bp in length
*    29559   29560: gap of unknown length
*    30139   30140: contig of 580 bp in length
*    30219   30220: gap of unknown length
*    30861        : contig of 642 bp in length
```

```
*      30862: gap of unknown length
*      30941: contig of 1395 bp in length
*      32336: gap of unknown length
*      32416: contig of 438 bp in length
*      32854: gap of unknown length
*      32934: contig of 1187 bp in length
*      34121: gap of unknown length
*      34201: contig of 732 bp in length
*      34933: gap of unknown length
*      35013: contig of 804 bp in length
*      35817: gap of unknown length
*      35897: contig of 750 bp in length
*      36647: gap of unknown length
*      36727: contig of 1083 bp in length
*      37810: gap of unknown length
*      37890: contig of 2013 bp in length
*      39903: gap of unknown length
*      39983: contig of 1116 bp in length
*      41099: gap of unknown length
*      41179: contig of 1490 bp in length
*      42669: gap of unknown length
*      42749: contig of 1387 bp in length
*      44136: gap of unknown length
*      44216: contig of 1203 bp in length
*      45419: gap of unknown length
*      45499: contig of 737 bp in length
*      46236: gap of unknown length
*      46316: contig of 1664 bp in length
*      47980: gap of unknown length
*      48060: contig of 1413 bp in length
*      49473: gap of unknown length
*      49553: contig of 1084 bp in length
*      50637: gap of unknown length
*      50717: contig of 1699 bp in length
*      52416: gap of unknown length
*      52496: contig of 871 bp in length
*      53367: gap of unknown length
*      53447: contig of 1261 bp in length
*      54708: gap of unknown length
*      54788: contig of 1549 bp in length
*      56337: gap of unknown length
*      56417: contig of 1475 bp in length
*      57892: gap of unknown length
*      57972: contig of 1099 bp in length
*      59071: gap of unknown length
*      59151: contig of 1848 bp in length
*      60999: gap of unknown length
*      61079: contig of 1677 bp in length
*      62756: gap of unknown length
*      62836: contig of 1211 bp in length
*      64047: gap of unknown length
*      64127: contig of 1210 bp in length
*      65337: gap of unknown length
*      65417: contig of 1613 bp in length
*      67030: gap of unknown length
*      67110: contig of 1342 bp in length
*      68452: gap of unknown length
*      68532: contig of 1937 bp in length
*      70469: gap of unknown length
*      70549: contig of 1121 bp in length
*      71670: gap of unknown length
*      71750: contig of 1265 bp in length
*      73015: gap of unknown length
*      73095: contig of 1007 bp in length
*      74102: gap of unknown length
*      74182: contig of 1359 bp in length
*      75541: gap of unknown length
*      75621: contig of 1479 bp in length
*      77100: gap of unknown length
*      77180: contig of 2680 bp in length
*      79860: gap of unknown length
*      79940: contig of 767 bp in length
*      80707: gap of unknown length
*      80787: contig of 2541 bp in length
*      83328: gap of unknown length
*      83408: contig of 2491 bp in length
*      85899: gap of unknown length
*      85979: contig of 2122 bp in length
*      88101: gap of unknown length
*      88181: contig of 2163 bp in length
*      90344: gap of unknown length
*      90424: contig of 3984 bp in length
*      94408: gap of unknown length
*      94488: contig of 4754 bp in length
*      99242: gap of unknown length
*      99322: contig of 597 bp in length
```

alignment_scores:
```
     Quality:  40.00              Length:          9
       Ratio:  4.444                Gaps:          0
Percent Similarity: 100.000   Percent Identity: 88.889
```

alignment_block:
US-08-653-294-11 x AC008327/rev

Align seg 1/1 to reverse of: AC008327   from: 1 to: 185469

```
     1 TyrArgLeuIleArgLeuAsnGlu 9
       ::|||||||||||||||||||||||
161633 TTCCGGTTACTTATTCGGTTGAACGAG 161607
```

seq_name: gb_htg7:AC017912

seq_documentation_block:
```
LOCUS       AC017912     49261 bp    DNA        HTG        09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017912
VERSION     AC017912.1  GI:6553278
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 49261)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212607 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..49261
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     14872 a    9975 c    9922 g    14492 t
ORIGIN
```

alignment_scores:
```
     Quality:  39.00              Length:         10
       Ratio:  3.900                Gaps:          0
Percent Similarity: 100.000   Percent Identity: 70.000
```

alignment_block:
US-08-653-294-11 x AC017912/rev

Align seg 1/1 to reverse of: AC017912   from: 1 to: 49261

```
     1 TyrArgLeuIleArgLeuAsnGluArg 10
       ||:::||||||||||||||:::|||:|||
```

8116 TATAAACTCTTGATCACAGGATCAATAGCCGA 8087

seq_name: gb_pr2:AC002465

seq_documentation_block:
LOCUS   AC002465   155881 bp   DNA   PRI   20-AUG-1997
DEFINITION Human BAC clone RG343P13 from 7q31, complete sequence.
ACCESSION   AC002465
VERSION   AC002465.1 GI:2337862
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 155881)
  AUTHORS   Connell,M and Cordes,M.
  TITLE   The sequence of H. sapiens BAC clone RG343P13
  JOURNAL   Unpublished (1997)
REFERENCE   2 (bases 1 to 155881)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from a release of the human BAC library. The library
            contains cloned DNA from human sperm. See: Shizuya et al., Proc.
            Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics
            34:213-8 (1996). The clone is available from Research Genetics,
            Inc. (http://www.resgen.com).
            VECTOR: pBeloBAC11
            Selection: chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RG343P13;
            actual end is at 155881 of RG343P13. The orientation of this clone
            is unknown.

            This clone contains STS's sWSS847 (NID:g1916380), sWSS846
            (NID:g1916379).
FEATURES   Location/Qualifiers
     source   1..155881
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RG343P13"

     repeat_region   /clone_lib="CITB-HS-A"
                     /map="7q31"
     repeat_region   complement(1528..1820)
                     /rpt_family="ALU"
     repeat_region   7349..7634
                     /rpt_family="ALU"
     repeat_region   7988..8279
                     /rpt_family="ALU"
     repeat_region   complement(9350..9639)
                     /rpt_family="ALU"
     repeat_region   complement(12067..12358)
                     /rpt_family="ALU"
     repeat_region   13275..13399
                     /rpt_family="ALU"
     repeat_region   complement(14341..14480)
                     /rpt_family="ALU"
     repeat_region   15092..15381
                     /rpt_family="ALU"
     repeat_region   complement(17529..17751)
                     /rpt_family="ALU"
     misc_feature   complement(17893..18003)
                     /note="match to human EST T99653 (NID:g749390) ye67a02.r1"
     misc_feature   complement(17893..18004)
                     /note="match to human EST W78848 (NID:g1389395)
                     zh51g03.r1"
     gene   complement(<17894..43271)
                     /gene="WNT2"
     CDS   complement(join(<17894..18158,35353..35630,40849..41075,
                     43189..43271))
                     /gene="WNT2"
                     /note="match to human mRNA X07876 (NID:g33970) and WNT2
                     protein P09544 (PID:g139750)"
                     /codon_start=1
                     /product="secreted growth factor"
                     /protein_id="AAB67043.1"
                     /db_xref="GI:2337863"
                     /translation="MNAPLGGIWLWLPLLTWLTPEVNSSWWYMRATGSSRVMCDNV
                     PGLVSSQRQLCHRHPDVMRAISQGVAEWTAECQHFRQHRWNCNTLDRDHSLFGRVLL
                     RSSRESAFVYAISSAGVVFAITRACSQGEVKSCSCDPKKMGSAKDSKGIFDWGGCSDN
                     IDYGIKFARAFVDAKERGKDARALMNLHNNRAGRKAVKRFLKQECKCHGVSGSCTLR
                     TCWLAMADFRKTGDYLWRKYNGAIQVVMNQDGTGFTVANERFKKPTKNDLVYFENSPD
                     YCIRDREA"
     repeat_region   complement(19783..20079)
                     /rpt_family="ALU"
     repeat_region   complement(20479..20766)
                     /rpt_family="ALU"
     repeat_region   20803..21090
                     /rpt_family="ALU"
     repeat_region   complement(22097..22377)
                     /rpt_family="ALU"
     repeat_region   22508..22606
                     /rpt_family="L1"
     repeat_region   complement(24726..25017)
                     /rpt_family="ALU"
     repeat_region   complement(25953..26002)
                     /rpt_family="ALU"
     repeat_region   complement(26028..26117)
                     /rpt_family="ALU"
     repeat_region   complement(26429..26477)
                     /rpt_family="L1"
     repeat_region   complement(26819..26845)
                     /rpt_family="L1"
     repeat_region   29184..29475
                     /rpt_family="L1"
     repeat_region   complement(30218..30369)
                     /rpt_family="ALU"
     repeat_region   30565..30874
                     /rpt_family="ALU"
     repeat_region   31884..31921
                     /rpt_family="L1"
     repeat_region   complement(32427..32453)
                     /rpt_family="L1"
     repeat_region   complement(32454..32587)

repeat_region   /rpt_family="ALU"
                 complement(32617..32681)
misc_feature   /rpt_family="ALU"
                 35367..35529
                 /gene="WNT2"
                 /note="match to human EST T29432 (NID:g611530)"
repeat_region   complement(38854..39080)
                 /rpt_family="ALU"
misc_feature   43070..45031
                 /note="CpG_island (%GC=64.4, o/e=0.70, #CpGs=157)"
repeat_region   47242..47360
                 /rpt_family="ALU"
repeat_region   complement(48223..48514)
                 /rpt_family="ALU"
repeat_region   48717..48995
                 /rpt_family="ALU"
repeat_region   complement(49005..49032)
                 /rpt_family="L1"
repeat_region   51231..51772
                 /rpt_family="L1"
repeat_region   52336..52494
                 /rpt_family="L1"
repeat_region   52528..52987
                 /rpt_family="L1"
repeat_region   54271..54366
                 /rpt_family="L1"
repeat_region   55051..55074
                 /rpt_family="L1"
repeat_region   55654..55791
                 /rpt_family="L1"
repeat_region   55902..55935
                 /rpt_family="L1"
repeat_region   56087..56189
                 /rpt_family="L1"
repeat_region   complement(56256..56548)
                 /rpt_family="ALU"
repeat_region   56655..56695
                 /rpt_family="L1"
repeat_region   56730..56802
                 /rpt_family="L1"
repeat_region   56818..56938
                 /rpt_family="ALU"
repeat_region   57144..57227
                 /rpt_family="L1"
repeat_region   complement(59426..59717)
                 /rpt_family="ALU"
repeat_region   complement(60815..61106)
                 /rpt_family="ALU"
repeat_region   complement(61569..61869)
                 /rpt_family="ALU"
repeat_region   complement(63011..67613)
                 /rpt_family="L1"
repeat_region   63980..64398
                 /rpt_family="L1"
repeat_region   67964..68015
                 /rpt_family="MER"
repeat_region   68059..69083
                 /rpt_family="MER"
repeat_region   72312..72350

alignment_scores:
        Quality: 39.00   Length: 10
        Ratio: 4.333   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AC002465

Align seg 1/1 to: AC002465 from: 1 to: 155881

        1 TyrArgLeuIleArgLeuAsnGluArg 10
          |||||||| ::||:||:||||||||
    15573 TATAGACTTTCTGTGAGAGTCAATGAGAGA 15602

---

seq_name: gb_htg5:AC015280

seq_documentation_block:
LOCUS   AC015280   159468 bp   DNA   HTG   16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015280
VERSION   AC015280.1 GI:6436055
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 159468)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10213506 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES   Location/Qualifiers
     source   1..159468
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT   47335 a   32541 c   32291 g   47301 t
ORIGIN

alignment_scores:
        Quality: 39.00   Length: 10
        Ratio: 3.900   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AC015280

Align seg 1/1 to: AC015280 from: 1 to: 159468

        1 TyrArgLeuIleArgLeuArgAsnGluArg 10
          ||:::|||||||||||||:::|||:|||
    44309 TATAAATATTGATCAGGATCAATAGCCGA 44338

seq_name: gb_htg4:AC007054

seq_documentation_block:
LOCUS   AC007054   171979 bp   DNA   HTG   13-OCT-1999
DEFINITION Drosophila melanogaster chromosome 2 clone BACR45O18 (D527) RPCI-98
            45.O.18 map 41E-41E strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 13 unordered pieces.
ACCESSION   AC007054
VERSION   AC007054.22 GI:6041715
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 171979)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blaze,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka.L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi.M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran.L.L. and
            Rubin,G.M.

```
us-08-653-294-11.rge
```

TITLE	Sequencing of Drosophila melanogaster
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 171979)
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT	On Oct 15, 1999 this sequence version replaced gi:5922045.

```
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*            651: contig of 651 bp in length
*            731: gap of unknown length
*            652: contig of 678 bp in length
*           1409: gap of unknown length
*           1410: contig of 95416 bp in length
*           1489: gap of unknown length
*           1840: contig of 351 bp in length
*           1490: gap of unknown length
*           1841: gap of unknown length
*           1920: gap of unknown length
*           1921: contig of 506 bp in length
*           2426: contig of 69242 bp in length
*           2427: gap of unknown length
*           2506: gap of unknown length
*           2507: gap of unknown length
*          71748: contig of 69242 bp in length
*          71749: gap of unknown length
*          71828: gap of unknown length
*          71829: contig of 95416 bp in length
*         167244: gap of unknown length
*         167245: contig of 618 bp in length
*         167324: gap of unknown length
*         167325: contig of 790 bp in length
*         167942: gap of unknown length
*         167943: gap of unknown length
*         168022: gap of unknown length
*         168023: contig of 652 bp in length
*         168812: gap of unknown length
*         168813: contig of 272 bp in length
*         168892: gap of unknown length
*         168893: gap of unknown length
*         169544: gap of unknown length
*         169545: contig of 338 bp in length
*         169624: gap of unknown length
*         169625: contig of 1021 bp in length
*         169896: gap of unknown length
*         169897: contig of 484 bp in length
*         169976: gap of unknown length
*         170314: gap of unknown length
*         170315: gap of unknown length
*         170394: gap of unknown length
*         170395: contig of
*         171415: gap of unknown length
*         171416: contig of
*         171495: gap of unknown length
*         171496: contig of
*         171979:
```

FEATURES	Location/Qualifiers
source	1..171979 /organism="Drosophila melanogaster" /strain="y; cn bw sp" /db_xref="taxon:7227" /chromosome="2" /map="41E-41E" /clone="BACR45O18 (D527) RPCI-98 45.O.18" /clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACe3.6"

```
BASE COUNT   50278 a  35745 c  35444 g  49502 t   1010 others
ORIGIN
```

alignment_scores:
```
Quality:    39.00              Length:     10
Ratio:       3.900              Gaps:        0
Percent Similarity: 100.000    Percent Identity: 70.000
```

```
alignment_block:
US-08-653-294-11 x AC007054/rev

Align seg 1/1 to reverse of: AC007054 from: 1 to: 171979

     1 TyrArgLeuIleArgLeuAsnGluArg 10
       |||:::|||||||||:::||:::||||
128523 TATAAACTCTTGATCAGGATCAATAGCCGA 128494

seq_name: gb_htg4:AC006467
```

seq_documentation_block:	
LOCUS	AC006467 175695 bp DNA HTG 27-OCT-1999
DEFINITION	Drosophila melanogaster chromosome 2 clone BACR03L08 (D532) RPCI-98 03.L.8 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION	AC006467
VERSION	AC006467.11 GI:6136329
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 175695)
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	Sequencing of Drosophila melanogaster
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 175695)
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT	On Oct 27, 1999 this sequence version replaced gi:5670661.

```
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.berkeley.edu). or send email
to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*              1:  550: contig of 550 bp in length
*            551:  630: gap of unknown length
*            631: 1229: contig of 599 bp in length
*           1230: 1309: gap of unknown length
*           1310: 1869: contig of 560 bp in length
*           1870: 1949: gap of unknown length
*           1950: 3022: contig of 1073 bp in length
*           3023: 3102: gap of unknown length
*           3103: 7823: contig of 4721 bp in length
```

```
*    7824      7903: gap of unknown length
*    7904    173453: contig of 165550 bp in length
*  173454    173533: gap of unknown length
*  173534    174264: contig of 731 bp in length
*  174265    174344: gap of unknown length
*  174345    174802: contig of 458 bp in length
*  174803    174882: gap of unknown length
*  174883    175695: contig of 813 bp in length.

FEATURES             Location/Qualifiers
     source          1..175695
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="40A-40C"
                     /clone="BACR03L08 (D532) RPCI-98 03.L.8"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"

BASE COUNT     51759 a    35773 c    35503 g    51919 t      641 others
ORIGIN

alignment_scores:
       Quality:     39.00          Length:       10
         Ratio:      3.900           Gaps:        0
Percent Similarity: 100.000    Percent Identity:  70.000

alignment_block:
US-08-653-294-11 x AC006467    ..

Align seg 1/1 to: AC006467    from:  1   to: 175695

      1 TyrArgLeuIleArgLeuAsnGluArg 10
        |||::|||||||||||||||||:::|||
  48184 TATAAATTATTGATCAGGATCAATAGCCGA 48213

seq_name: gb_htg3:AC008682

seq_documentation_block:
LOCUS       AC008682    216649 bp    DNA                 HTG       03-AUG-1999
DEFINITION  Homo sapiens chromosome 5 clone CIT978SKB_54G2, *** SEQUENCING IN
            PROGRESS ***, 63 unordered pieces.
ACCESSION   AC008682
VERSION     AC008682.1  GI:5685896
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216649)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216649)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 63 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      637: contig of 637 bp in length
            *      638           gap of unknown length
            *             1631: contig of 994 bp in length
            *     1632           gap of unknown length
            *             2553: contig of 922 bp in length

*    2554      3319: gap of unknown length
*    3320      3994: contig of 675 bp in length
*    3995      4648: gap of unknown length
*    4649      5512: contig of 864 bp in length
*    5513      6208: gap of unknown length
*    6209      6922: contig of 714 bp in length
*    6923      8025: gap of unknown length
*    8026      9274: contig of 1249 bp in length
*    9275     10498: gap of unknown length
*   10499     11634: contig of 1136 bp in length
*   11635     12332: gap of unknown length
*   12333     13945: contig of 1613 bp in length
*   13946     15278: gap of unknown length
*   15279     16574: contig of 1296 bp in length
*   16575     17988: gap of unknown length
*   17989     19075: contig of 1087 bp in length
*   19076     20518: gap of unknown length
*   20519     21769: contig of 1251 bp in length
*   21770     22593: gap of unknown length
*   22594     23584: contig of 991 bp in length
*   23585     24947: gap of unknown length
*   24948     26399: contig of 1452 bp in length
*   26400     27550: gap of unknown length
*   27551     28267: contig of 717 bp in length
*   28268     29109: gap of unknown length
*   29110     30110: contig of 1001 bp in length
*   30111     31414: gap of unknown length
*   31415     32809: contig of 1395 bp in length
*   32810     34507: gap of unknown length
*   34508     36466: contig of 1959 bp in length
*   36467     38262: gap of unknown length
*   38263     39740: contig of 1478 bp in length
*   39741     40973: gap of unknown length
*   40974     43125: contig of 2152 bp in length
*   43126     46386: gap of unknown length
*   46387     48948: contig of 2562 bp in length
```

```
*       48949     50932:  contig of 1984 bp in length
*       50933     54161:  gap of unknown length
*       54162     56414:  contig of 3229 bp in length
*       56415     60612:  gap of unknown length
*       60613     62409:  contig of 4198 bp in length
*       62410     66885:  gap of unknown length
*       66886     70123:  contig of 1797 bp in length
*       70124     74534:  gap of unknown length
*       74535     78215:  contig of 4476 bp in length
*       78216     82302:  gap of unknown length
*       82303     85613:  contig of 3238 bp in length
*       85614     87774:  gap of unknown length
*       87775     91685:  contig of 4411 bp in length
*       91686     99387:  gap of unknown length
*       99388    103660:  contig of 3681 bp in length
*      103661    109950:  gap of unknown length
*      109951    119790:  contig of 4087 bp in length
*      119791    126523:  gap of unknown length
*      126524    131530:  contig of 3311 bp in length
*      131531    138930:  gap of unknown length
*      138931    149780:  contig of 2161 bp in length
*      149781    165181:  gap of unknown length
*      165182    182085:  contig of 3911 bp in length
*      182086    216649:  gap of unknown length
                          contig of 7702 bp in length
                          gap of unknown length
                          contig of 4273 bp in length
                          gap of unknown length
                          contig of 6290 bp in length
                          gap of unknown length
                          contig of 9840 bp in length
                          gap of unknown length
                          contig of 6733 bp in length
                          gap of unknown length
                          contig of 5007 bp in length
                          gap of unknown length
                          contig of 7400 bp in length
                          gap of unknown length
                          contig of 10850 bp in length
                          gap of unknown length
                          contig of 15401 bp in length
                          gap of unknown length
                          contig of 16904 bp in length
                          gap of unknown length
                          contig of 34564 bp in length.
FEATURES             Location/Qualifiers
     source          1..216649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CIT978SKB_54G2"
BASE COUNT      66797 a   42466 c   41332 g   66015 t      39 others
ORIGIN
```

alignment_scores:
     Quality: 39.00   Length: 10
     Ratio: 3.900   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AC008682/rev

Align seg 1/1 to reverse of: AC008682 from: 1 to: 216649

   1 TyrArgLeuIleArgLeuAsnGluArg 10
     |||||::::||||||||||::::||||:|||
5329 TATAGGGTATTAATAAGGATAAATAATCGG 5300

OM of: US-08-653-294-11 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000 1:27 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgnl_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPEXT=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT-0.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=-1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-08-653-294-11
Query length: 10
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096542
Search time (sec): 590.520000

N_Geneseq_36:V73474	-		32.00	97.61	775.93	1769	! Mouse G3BP cDNA. New ubiqui
N_Geneseq_36:Q77686	+		32.00	96.17	933.97	2086	! Temp. sensitive autolysing
N_Geneseq_36:T31725	-		32.00	95.99	955.65	2129	! Human GAP-SH3 domain bindin
N_Geneseq_36:N91438	-		32.00	92.80	1.4e+03	3063	! DNA encoding glycine rich P

seq_name: N_Geneseq_36:T29774

seq_documentation_block:
ID T29774 standard; DNA; 2040 BP.
AC T29774;
DT 19-NOV-1996 (first entry)
DE Bacillus thuringiensis ssp. israelensis CryIVD protein DNA.
KW CryIVD; toxic protein; crystal toxin; expression construct;
KW transformed cyanobacteria; phycocyanin beta; cpcB; promoter;
KW insecticide; dipteran larvae; mosquito; blackfly; ss.
OS Bacillus thuringiensis.
FH Key   Location/Qualifiers
FT cds   41..1972
FT   /*tag= a
PN US5518897-A.
PD 21-MAY-1996.
PF 04-MAY-1992; 877876.
PR 04-MAY-1992; US-877876.
PR 28-JAN-1994; US-188581.
PA (UYME-) UNIV MEMPHIS STATE.
PI Murphy RC, Stevens SE;
DR WPI: 96-259063/26.
DR P-PSDB; R97735.
PT New DNA construct for expressing cryIV D protein in cyanobacteria -
PT under control of a phycocyanin beta promoter, useful for control of
PT dipteran larvae in water
PS Example 1; Columns 9-14; 20pp; English.
CC The present sequence encodes the B. thuringiensis ssp. israelensis
CC CryIVD toxic protein, which was used in the prepn. of a claimed DNA
CC construct for the expression of CryIVD in cyanobacteria, comprising
CC the present sequence under the control of phycocyanin beta (cpcB)
CC promoter. Cyanobacteria (which may be adapted for growth in fresh
CC or brackish water) transformed with the construct can be used
CC as insecticides for controlling dipteran larvae, esp. those of
CC mosquitoes and blackflies, that live in water. The percentage of
CC Culex pipens (mosquito) larvae surviving after 4 days with
CC cyanobacteria transformed with the claimed DNA construct as their
CC only food source was 51 %, compared to 94 % for those fed with
CC cyanobacteria transformed with an empty plasmid. In the
CC cyanobacteria, CryIVD is efficiently expressed under the control of
CC the strong cpcB promoter, even though the CryIVD gene contains 19
CC AUA which are generally poorly translated (if at all) in
CC cyanobacteria.
SQ Sequence 2040 BP; 716 A; 315 C; 373 G; 636 T;

alignment_scores:
                        Quality: 37.00   Length: 9
                        Ratio: 4.111   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x T29774/rev

Align seg 1/1 to reverse of: T29774 from: 1 to: 2040

        1 TyrArgLeuLeuArgLeuAsnGlu 9
          :::||||||||||||::||||||||
      542 TTCCGTCTTTTAATAAGTAAATGAA 516

seq_name: N_Geneseq_36:N90712

seq_documentation_block:
ID N90712 standard; DNA; 2100 BP.
AC N90712;
DT 09-JAN-1990 (first entry)
DE cryD gene.
KW cryD protein; Bacillus thuringiensis; biopesticide.

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
N_Geneseq_36:T29774	-	37.00	115.12	82.10	2040	! Bacillus thuringiensis ssp. is
N_Geneseq_36:N90712	-	37.00	114.87	84.82	2100	! cryD gene. Bacillus thuringien
N_Geneseq_36:Q14809	-	37.00	112.34	117.32	2802	! Btm PG14 72kDa Cry insecticida
N_Geneseq_36:Q14810	-	37.00	112.04	121.99	2901	! Btm PG14 72kDa Cry insecticida
N_Geneseq_36:V84560	+	35.00	108.51	191.66	1842	! Human secreted protein gene 15
N_Geneseq_36:X20531	-	35.00	102.03	440.16	3858	! Polynucleotide sequence from t
N_Geneseq_36:V65691	+	35.00	120.34	621.49	5243	! Fowlpox virus vector DNA fragm
N_Geneseq_36:T20013	-	34.00	119.34	41.34	307	! Human gene signature HUMGS01151
N_Geneseq_36:V77268	-	34.00	118.29	54.74	394	! Staphylococcus aureus contig SE
N_Geneseq_36:X18185	-	34.00	111.97	123.10	810	! S. aureus gidB coding sequence.
N_Geneseq_36:X18186	-	34.00	111.84	138.59	900	! S. aureus gidB ORF sequence. Ne
N_Geneseq_36:V82078	+	34.00	107.82	209.57	1300	! DNA encoding a partial gidAl F
N_Geneseq_36:V74973	-	34.00	107.18	227.42	1398	! Staphylococcus aureus contig S
N_Geneseq_36:Q20995	-	34.00	106.22	257.27	1560	! EHV-4 gC gene. Nucleic acid se
N_Geneseq_36:V82077	-	34.00	103.48	365.56	2132	! DNA encoding a gidAl protein.
N_Geneseq_36:V07921	-	34.00	102.67	405.51	2338	! Helicobacter pylori 76 kDa pol
N_Geneseq_36:V07916	-	34.00	102.34	423.30	2429	! Helicobacter pylori 76 kDa pol
N_Geneseq_36:V74500	-	34.00	96.30	917.45	4832	! Staphylococcus aureus contig S
N_Geneseq_36:X20560	-	34.00	91.39	1.7e+03	8467	! Polynucleotide sequence from t
N_Geneseq_36:V74420	-	34.00	84.85	4.0e+03	17846	! Polynucleotide sequence from t
N_Geneseq_36:V48249	-	33.00	103.71	354.70	1353	! Nucleotide sequence encoding s
N_Geneseq_36:V74701	-	33.00	101.27	485.32	1788	! Staphylococcus aureus contig s
N_Geneseq_36:V74662	-	33.00	101.13	493.87	1816	! Staphylococcus aureus contig s
N_Geneseq_36:X13313	+	33.00	98.90	658.07	2344	! Enterococcus faecalis genome c
N_Geneseq_36:T42751	+	33.00	89.81	2.1e+03	6608	! Chicken CHD-1A gene. Avian chr
N_Geneseq_36:V20767	+	33.00	86.26	3.3e+03	9898	! Human OCIF genome DNA-2. Inhib
N_Geneseq_36:T33183	+	32.50	86.01	3.4e+03	10190	! Fragment of human OCIF genomi
N_Geneseq_36:X13065	-	33.00	75.77	1.3e+04	32768	! Enterococcus faecalis genome
N_Geneseq_36:X13095	-	33.00	79.92	7.5e+03	16484	! Enterococcus faecalis genome
N_Geneseq_36:Q39810	+	32.00	112.00	122.63	343	! Expressed Sequence Tag human ge
N_Geneseq_36:Q59222	+	32.00	112.00	122.63	343	! Human brain Expressed Sequence
N_Geneseq_36:T20671	-	32.00	109.73	163.93	444	! Human gene signature HUMGS01887
N_Geneseq_36:Q99805	+	32.00	104.24	331.74	831	! Thaumatin like gene PR-5mz. New
N_Geneseq_36:Q70084	-	32.00	103.79	351.11	874	! Bacillus thuringiensis transcri
N_Geneseq_36:T84095	+	32.00	102.34	423.26	1032	! DNA encoding a staphylococcus
N_Geneseq_36:V75169	+	32.00	101.19	490.23	1176	! Staphylococcus aureus contig s
N_Geneseq_36:T33136	-	32.00	99.76	588.77	1384	! Broccoli ACC synthase genomic
N_Geneseq_36:Q13581	-	32.00	99.64	598.35	1404	! A.altocetigenes membrane-bound
N_Geneseq_36:Q20384	-	32.00	99.64	598.35	1404	! ADH complex structural gene (4
N_Geneseq_36:T59978	+	32.00	98.78	668.28	1549	! 3' untranslated region of the
N_Geneseq_36:X02018	-	32.00	98.78	668.28	1549	! D. melanogaster tipE+ 4Kb clon

OS Bacillus thuringiensis var.israelesis.
PN WO9907605-A.
PD 24-AUG-1989.
PF 17-FEB-1989; U00663.
PR 19-FEB-1988; US-158176.
PA (ECOG) Ecogen Inc.
PI Donovan WP;
DR WPI: 89-263682/36.
DR P-PSDB; P91462.
PT Bacillus thuringiensis var israelansis cry D toxin gene and proteins
PT - used for producing insecticide compsns.active against Dipteran species.
PS Claim 1; fig 2; 58pp; English.
CC cryD gene is inserted into plasmid and used to transform a microorganism.
CC The 67kD protein encoded by the gene has insecticidal activity against
CC dipteran larvae.
SQ Sequence 2100 BP; 746 A; 316 C; 378 G; 660 T;

alignment_scores:
  Quality: 37.00   Length: 9
  Ratio: 4.111   Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x N90712/rev ..
Align seg 1/1 to reverse of: N90712 from: 1 to: 2100

  1 TyrArgLeuIleArgLeuAsnGlu 9
   ::!|||||||||||::!||||||||
 542 TTCCGTCTTTAATAAAGTAAATGAA 516

seq_name: N_Geneseq_36:Q14809

seq_documentation_block:
ID Q14809 standard; DNA; 2802 BP.
AC Q14809;
DT 10-FEB-1992 (first entry)
DE Btm PG14 72kDa Cry insecticidal protein/25kDa Cyt A fusion gene.
KW chimeric; fusion protein; insecticide; Lepidoptera larvae;
KW midgut targetting; bacterial endotoxin; ss.
OS Bacillus thuringiensis subspecies morrisoni.
PN WO9117254-A.
PD 14-NOV-1991.
PF 02-MAY-1991; U03008.
PR 03-MAY-1990; US-518575.
PA (REGC) UNIV OF CALIFORNIA.
PI Sivasubramanian N, Federici A;
DR WPI: 91-353775/48.
PT Extending host range or toxicity of insecticidal proteins - using
PT protein capable of binding to gut epithelium of insects
PS Claim 34; Fig 21; 61pp; English.
CC This fusion gene comprises sequences isolated from the PG14 strain
CC of B.thuringiensis subsp. morrisoni. The cytA gene encodes a
CC cytolytic protein which has high affinity for the lipid portion of
CC cell membranes. After ingestion by insects, the 27 kDa Cyt A protein
CC is cleaved by midgut proteases to a relatively resistant core of 25
CC kDa. The Cyt A coding sequence is combined with the Cry toxin, also
CC from B.thuringiensis.
SQ Sequence 2802 BP; 984 A; 395 C; 544 G; 879 T;

alignment_scores:
  Quality: 37.00   Length: 9
  Ratio: 4.111   Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x Q14809/rev ..
Align seg 1/1 to reverse of: Q14809 from: 1 to: 2802

  1 TyrArgLeuIleArgLeuAsnGlu 9

   ::!|||||||||||::!||||||||
 542 TTCCGTCTTTAATAAAGTAAATGAA 516

seq_name: N_Geneseq_36:Q14810

seq_documentation_block:
ID Q14810 standard; DNA; 2901 BP.
AC Q14810;
DT 10-FEB-1992 (first entry)
DE Btm PG14 72kDa Cry insecticidal protein/27kDa Cyt A fusion gene.
KW chimeric; fusion protein; insecticide; Lepidoptera larvae;
KW midgut targetting; Cry A; bacterial endotoxin; ss.
OS Bacillus thuringiensis subspecies morrisoni.
PN WO9117254-A.
PD 14-NOV-1991.
PF 02-MAY-1991; U03008.
PR 03-MAY-1990; US-518575.
PA (REGC) UNIV OF CALIFORNIA.
PI Sivasubramanian N, Federici A;
DR WPI: 91-353775/48.
PT Extending host range or toxicity of insecticidal proteins - using
PT protein capable of binding to gut epithelium of insects
PS Claim 33; Fig 22; 61pp; English.
CC This fusion gene comprises sequences isolated from the PG14 strain
CC of B.thuringiensis subsp. morrisoni. The cytA gene encodes a
CC cytolytic protein which has high affinity for the lipid portion of
CC cell membranes. After ingestion by insects, the 27 kDa Cyt A protein
CC is cleaved by midgut proteases to a relatively resistant core of 25
CC kDa. The Cyt A coding sequence is combined with the Cry toxin, also
CC from B.thuringiensis.
SQ Sequence 2901 BP; 1023 A; 462 C; 503 G; 913 T;

alignment_scores:
  Quality: 37.00   Length: 9
  Ratio: 4.111   Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x Q14810/rev ..
Align seg 1/1 to reverse of: Q14810 from: 1 to: 2901

  1 TyrArgLeuIleArgLeuAsnGlu 9
   ::!|||||||||||::!||||||||
 542 TTCCGTCTTTAATAAAGTAAATGAA 516

seq_name: N_Geneseq_36:V84560

seq_documentation_block:
ID V84560 standard; DNA; 1842 BP.
AC V84560;
DT 01-MAR-1999 (first entry)
DE Human secreted protein gene 150 clone HMSKQ35.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9854963-A2.
PD 10-DEC-1998.
PF 04-JUN-1998; U11422.
PR 18-DEC-1997; US-070923.
PR 06-JUN-1997; US-048877.
PR 06-JUN-1997; US-048881.
PR 06-JUN-1997; US-048884.
PR 06-JUN-1997; US-048893.
PR 06-JUN-1997; US-048896.
PR 06-JUN-1997; US-048899.

PR 06-JUN-1997; US-048915.
PR 06-JUN-1997; US-048949.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048972.
PR 06-JUN-1997; US-049020.
PR 06-JUN-1997; US-049375.
PR 05-SEP-1997; US-057628.
PR 05-SEP-1997; US-057635.
PR 05-SEP-1997; US-057644.
PR 05-SEP-1997; US-057647.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057661.
PR 05-SEP-1997; US-057667.
PR 05-SEP-1997; US-057761.
PR 05-SEP-1997; US-057764.
PR 05-SEP-1997; US-057770.
PR 05-SEP-1997; US-057775.
PR 05-SEP-1997; US-057778.
PR 06-JUN-1997; US-048875.
PR 06-JUN-1997; US-048878.
PR 06-JUN-1997; US-048882.
PR 06-JUN-1997; US-048885.
PR 06-JUN-1997; US-048894.
PR 06-JUN-1997; US-048897.
PR 06-JUN-1997; US-048900.
PR 06-JUN-1997; US-048901.
PR 06-JUN-1997; US-048916.
PR 06-JUN-1997; US-048962.
PR 06-JUN-1997; US-048963.
PR 06-JUN-1997; US-048974.
PR 06-JUN-1997; US-049373.
PR 05-SEP-1997; US-057584.
PR 05-SEP-1997; US-057629.
PR 05-SEP-1997; US-057642.
PR 05-SEP-1997; US-057645.
PR 05-SEP-1997; US-057648.
PR 05-SEP-1997; US-057651.
PR 05-SEP-1997; US-057662.
PR 05-SEP-1997; US-057668.
PR 05-SEP-1997; US-057762.
PR 05-SEP-1997; US-057765.
PR 05-SEP-1997; US-057771.
PR 05-SEP-1997; US-057776.
PR 06-JUN-1997; US-048876.
PR 06-JUN-1997; US-048880.
PR 06-JUN-1997; US-048883.
PR 06-JUN-1997; US-048892.
PR 06-JUN-1997; US-048895.
PR 06-JUN-1997; US-048898.
PR 06-JUN-1997; US-048901.
PR 06-JUN-1997; US-048917.
PR 06-JUN-1997; US-048963.
PR 06-JUN-1997; US-048971.
PR 06-JUN-1997; US-049019.
PR 06-JUN-1997; US-049374.
PR 05-SEP-1997; US-057627.
PR 05-SEP-1997; US-057634.
PR 05-SEP-1997; US-057643.
PR 05-SEP-1997; US-057646.
PR 05-SEP-1997; US-057649.
PR 05-SEP-1997; US-057654.
PR 05-SEP-1997; US-057666.
PR 05-SEP-1997; US-057760.
PR 05-SEP-1997; US-057763.
PR 05-SEP-1997; US-057769.
PR 05-SEP-1997; US-057774.
PR 05-SEP-1997; US-057777.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI Li Y, Moore DA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR WPI; 99-059865/05.

DR P-PSDB; W88683, W89013, W89014, W89015, W89016.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PS disorders, immune diseases, inflammation or blood disorders
CC Claim 4; Page 412-413; 772pp; English.
CC The invention relates to nucleic acid sequences (V84411 to V84633)
CC encoding human secreted proteins (W88534 to W88756). The secreted protein
CC gene sequences are deposited with the ATCC under deposit numbers ATCC
CC 97979, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC diseases, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
SQ Sequence 1842 BP; 604 A; 322 C; 369 G; 533 T;

alignment_scores:
          Quality: 35.00   Length: 8
          Ratio: 4.375   Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000

alignment_block:
US-08-653-294-11 x V84560   .

Align seg 1/1 to: V84560 from: 1 to: 1842

    1 TyrArgLeuIleArgLeuAsn 8
      |||||||||||||::||::::||
 1560 TACAGATTACTACTACGAATGAAT 1583

seq_name: N_Geneseq_36:X20531

seq_documentation_block:
ID X20531 standard; DNA; 3858 BP.
AC X20531;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 368-370; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the

```
CC   production of biosynthetic products such as enzymes.
SQ   Sequence 3858 BP;   659 A;   777 C;   1306 G;   1116 T;

alignment_scores:
        Quality:   35.00      Length:    9
          Ratio:   4.375      Gaps:      0
Percent Similarity: 88.889   Percent Identity: 77.778

alignment_block:
US-08-653-294-11 x X20531

Align seg 1/1  to reverse of: X20531  from: 1  to: 3858

      2 ArgLeuIleArgLeuAsnGluArg 10
        |||:|||||||||||||||||||
   2281 CGGATCACCATTAGCTGAACGAGCGC 2255

seq_name: N_Geneseq_36:V65691

seq_documentation_block:
ID   V65691 standard: DNA: 5243 BP.
AC   V65691;
DT   25-JAN-1999 (first entry)
DE   Fowlpox Virus vector DNA fragment.
KW   vector; poxvirus; infection; treatment; prevention; copy number;
KW   promoter; vaccine; ds.
OS   Fowlpox virus.
PN   WO9844093-A1.
PD   08-OCT-1998.
PF   26-MAR-1998; J01358.
PR   28-MAR-1997; JP-094875.
PA   (JAPG ) NIPPON ZEON KK.
PI   Sato T;
DR   WPI: 98-557105/47.
PT   Fowl poxvirus-originated 5 kb DNA vector with hetrologous genes    - -
PT   useful as a vaccine, providing large copy number per cell, with
PT   superior preventive efficacy
PS   Claim 2; Page 36-39; 58pp; Japanese.
CC   This sequence is used in the construction of a new vector originated
CC   from fowlpox virus which can duplicate poxvirus in infected cells. The
CC   DNA vector can be used to produce vaccines for preventing and/or treating
CC   infection caused by poxvirus including orthopoxvirus and chicken
CC   poxvirus. The DNA vector is about 5 kb and is concomitant. It can grow in
CC   non-poxvirus-infective cells. It has a large copy number per cell, making
CC   the vaccine very effective.
SQ   Sequence 5243 BP;   1826 A;   882 C;   1026 G;   1508 T;

alignment_scores:
        Quality:   35.00      Length:    9
          Ratio:   4.375      Gaps:      0
Percent Similarity: 88.889   Percent Identity: 88.889

alignment_block:
US-08-653-294-11 x V65691

Align seg 1/1  to: V65691  from: 1  to: 5243

      1 TyrArgLeuIleArgLeuAsnGlu 9
        ||||||||::|||:::|||  ||||
   3138 TATCGACTACTCGTCAAATACAACGAA 3164

seq_name: N_Geneseq_36:T20013

seq_documentation_block:
ID   T20013 standard: cDNA to mRNA; 307 BP.
AC   T20013;
DT   17-JUL-1996 (first entry)
DE   Human gene signature HUMGS01151.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
```

```
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI: 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Pages 536-537; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence 307 BP;   93 A;   49 C;   64 G;   101 T;

alignment_scores:
        Quality:   34.00      Length:    9
          Ratio:   3.778      Gaps:      0
Percent Similarity: 100.000   Percent Identity: 55.556

alignment_block:
US-08-653-294-11 x T20013

Align seg 1/1  to: T20013  from: 1  to: 307

      1 TyrArgLeuIleArgLeuAsnGlu 9
        ||||||:::|||:||:::|||::||::
    258 TACAGAATATTAGTAAGATTTAATCAA 284

seq_name: N_Geneseq_36:V77268

seq_documentation_block:
ID   V77268 standard: DNA; 394 BP.
AC   V77268;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #2957.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI: 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus    -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 2383; 327pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
```

CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 394 BP; 109 A; 85 C; 71 G; 126 T;

alignment_scores:
          Quality: 34.00   Length: 10
          Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-11 x V77268   ..

Align seg 1/1 to: V77268   from: 1 to: 394

     1 TyrArgLeuLeuIleArgLeuAsnGluArg 10
       ||::: ::|||||||||||||:|||
    88 TACAAATACATCATTCGACTTAATCGCAGG 117

seq_name: N_Geneseq_36:X18185

seq_documentation_block:
ID   X18185 standard; DNA; 810 BP.
AC   X18185;
DT   07-MAY-1999   (first entry)
DE   S. aureus gidB coding sequence.
KW   GidB; mutation detection;bacteriostatic; bacteriocidal compound;
KW   microbial infection; osteomyelitis; septic arthritis; gene therapy;
KW   septic thrombophlebitis; acute bacterial endocarditis; bacteraemia;
KW   cancer; ds.
OS   Staphylococcus aureus.
PN   EP-892055-A2.
PD   20-JAN-1999.
PF   30-JUN-1998; 305175.
PR   12-JUN-1998; US-097072.
PR   01-JUL-1997; US-886638.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
PI   Burnham MKR, Kallender H, Palmer LM, Ward J;
DR   WPI; 99-083572/08.
DR   P-PSDB; W74405.
PT   New Staphylococcus aureus GidB polypeptides and polynucleotides. -
PT   useful as diagnostic reagents and for prevention and treatment of
PT   Staphylococcus infections
PS   Claim 2; Page 21; 39pp; English.
CC   This sequence encodes the Staphylococcus aureus gidB protein of the
CC   invention. GidB polynucleotides and polypeptides are useful for
CC   diagnosing susceptibility to diseases by detecting mutations or
CC   polymorphisms in the GidB gene or analysing for the presence of amount of
CC   GidB polypeptide expressed in a patient sample. GidB PCR probes are
CC   useful for diagnosing diseases, and can characterise the response of the
CC   infectious organism to drugs. GidB polypeptides and polynucleotides are
CC   also useful for screening for antagonists, agonists and drugs against
CC   infectious micro-organisms. GidB agonists and antagonists are
CC   bacteriostatic and bacteriocidal compounds which can be used in treatment
CC   to enhance or block GidB activity, therefore treating diseases caused by
CC   microbial infection, especially S. aureus diseases including
CC   osteomyelitis, septic arthritis, septic thrombophlebitis, acute bacterial
CC   endocarditis and bacteraemia in cancer patients. Epitopes of GidB
CC   polypeptides and polynucleotides are useful immunogens for producing

CC   anti-GidB antibodies for prevention of bacterial infections, and GidB
CC   polynucleotides can be used in genetic immunisation to prevent
CC   infections. GidB polypeptides, polynucleotides and their (ant)agonists
CC   can prevent adhesion of bacteria to matrix proteins, and are useful for
CC   use on wounds and body implants to prevent bacterial infection. GidB
CC   polypeptides and polynucleotides may also be used as reagents for
CC   differential screening methods.
SQ   Sequence 810 BP; 295 A; 113 C; 158 G; 244 T;

alignment_scores:
          Quality: 34.00   Length: 10
          Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-11 x X18185   ..

Align seg 1/1 to: X18185   from: 1 to: 810

     1 TyrArgLeuLeuIleArgLeuAsnGluArg 10
       |||||||||||||::::|||:::||
   138 TATCGTTTACTTGTTGAATGGAATGAAAAG 167

seq_name: N_Geneseq_36:X18186

seq_documentation_block:
ID   X18186 standard; DNA; 900 BP.
AC   X18186;
DT   07-MAY-1999   (first entry)
DE   S. aureus gidB ORF sequence.
KW   GidB; mutation detection; bacteriostatic; bacteriocidal compound;
KW   microbial infection; osteomyelitis; septic arthritis; gene therapy;
KW   septic thrombophlebitis; acute bacterial endocarditis; bacteraemia;
KW   cancer; ds.
OS   Staphylococcus aureus.
PN   EP-892055-A2.
PD   20-JAN-1999.
PF   30-JUN-1998; 305175.
PR   12-JUN-1998; US-097072.
PR   01-JUL-1997; US-886638.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
PI   Burnham MKR, Kallender H, Palmer LM, Ward J;
DR   WPI; 99-083572/08.
DR   P-PSDB; W74406.
PT   New Staphylococcus aureus GidB polypeptides and polynucleotides -
PT   useful as diagnostic reagents and for prevention and treatment of
PT   Staphylococcus infections
PS   Claim 2; Page 23; 39pp; English.
CC   This sequence encodes the Staphylococcus aureus gidB protein of the
CC   invention. GidB polynucleotides and polypeptides are useful for
CC   diagnosing susceptibility to diseases by detecting mutations or
CC   polymorphisms in the GidB gene or analysing for the presence of amount of
CC   GidB polypeptide expressed in a patient sample. GidB PCR probes are
CC   useful for diagnosing diseases, and can characterise the response of the
CC   infectious organism to drugs. GidB polypeptides and polynucleotides are
CC   also useful for screening for antagonists, agonists and drugs against
CC   infectious micro-organisms. GidB agonists and antagonists are
CC   bacteriostatic and bacteriocidal compounds which can be used in treatment
CC   to enhance or block GidB activity, therefore treating diseases caused by
CC   microbial infection, especially S. aureus diseases including
CC   osteomyelitis, septic arthritis, septic thrombophlebitis, acute bacterial
CC   endocarditis and bacteraemia in cancer patients. Epitopes of GidB
CC   polypeptides and polynucleotides are useful immunogens for producing
CC   anti-GidB antibodies for prevention of bacterial infection, and GidB
CC   polynucleotides can be used in genetic immunisation to prevent
CC   infections. GidB polypeptides, polynucleotides and their (ant)agonists
CC   can prevent adhesion of bacteria to matrix proteins, and are useful for
CC   use on wounds and body implants to prevent bacterial infection. GidB
CC   polypeptides and polynucleotides may also be used as regents for
CC   differential screening methods.
SQ   Sequence 900 BP; 338 A; 124 C; 165 G; 267 T;

alignment_scores:
    Quality: 34.00   Length: 10
        Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-11 x X18186
Align seg 1/1 to: X18186 from: 1 to: 900

    1 TyrArgLeuIleArgLeuAsnGluArg 10
      ||||||||||||::::: ||||||::
  156 TATCGTTTACTTGTTGAATGGAATGAAAAG 185

seq_name: N_Geneseq_36:V82078

seq_documentation_block:
ID V82078 standard; DNA; 1300 BP.
AC V82078;
DT 25-MAR-1999 (first entry)
DE DNA encoding a partial gidA1 protein.
KW GidA1; bacterial infection; meningitis; Helicobacter pylori infection;
KW cancer; ulcer; gastritis; antibacterial; in-dwelling device;
KW wound treatment; bacterial adhesion; matrix protein; ds.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT CDS   2. .1165
FT   /tag= a
FT   /transl_except= (pos: 848. .850, aa: Xaa)
FT   /product= gidA1
FT   /note= "xaa= unspecified amino acid"
PN EP-889129-A2.
PD 07-JAN-1999.
PF 30-JUN-1998; 305180.
PR 01-JUL-1997; US-052758.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Burnham M, Kallender H, Lenox AL, Palmer LM;
DR WPI; 99-062660/06.
DR P-PSDB; W89446.
PT New isolated gidA1 polypeptide from Staphylococcus aureus - used to
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT H. pylori and associated cancers, ulcers and gastritis
PS Claim 2; Page 5-6; 43pp; English.
CC The present sequence encodes a partial gidA1 protein of Staphylococcus
CC aureus. GidA1 proteins, nucleic acids and agonists are used to
CC treat conditions requiring increased activity or expression of gidA1,
CC while conditions (particularly bacterial infections) requiring
CC inhibition of gidA1 are treated by administering an antagonist,
CC inhibitory nucleic acid or competitive polypeptide. The products are
CC used to treat S. pneumoniae infection, particularly meningitis and
CC also Helicobacter pylori infections e.g. related cancers, ulcers and
CC gastritis. These antibacterial agents may also be used to treat
CC in-dwelling devices to prevent infection or generally as wound
CC treatments to prevent adhesion of bacteria to matrix proteins.
SQ Sequence 1300 BP; 477 A; 195 C; 262 G; 365 T;

alignment_scores:
    Quality: 34.00   Length: 10
        Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-11 x V82078
Align seg 1/1 to: V82078 from: 1 to: 1300

    1 TyrArgLeuIleArgLeuAsnGluArg 10
      ||||||||||||::::: ||||||::
 1249 TATCGTTTACTTGTTGAATGGAATGAAAAG 1278

seq_name: N_Geneseq_36:V74973

seq_documentation_block:
ID V74973 standard; DNA; 1398 BP.
AC V74973;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #662.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT misc_feature   1261. .1320
FT   /tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1586-1587; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 1398 BP; 461 A; 197 C; 186 G; 494 T;

alignment_scores:
    Quality: 34.00   Length: 10
        Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 50.000

alignment_block:
US-08-653-294-11 x V74973/rev
Align seg 1/1 to reverse of: V74973 from: 1 to: 1398

    1 TyrArgLeuIleArgLeuAsnGluArg 10
      |||||: :::::||||||||||::|||
 1064 TATAAGAAGATTGTCCGATTAAATCAGCGC 1035

seq_name: N_Geneseq_36:Q20995

seq_documentation_block:
ID Q20995 standard; DNA; 1560 BP.
AC Q20995;
DT 19-MAY-1992 (first entry)

DE EHV-4 gC gene.
KW Equine herpes virus-4; glycoprotein gC; antigenic; vaccine;
KW alphaherpesvirus; respiratory disease; cellular attatchment;
KW pathogenic; ss.
OS Equine herpesvirus-4.
FH Key   Location/Qualifiers
FT cds   52..1509
FT   /*tag= a
FT   /product= EHV-4_gC
PN WO9201057-A.
PD 23-JAN-1992.
PF 04-JUL-1991; G01091.
PR 06-JUL-1990; GB-014950.
PA (UNIU) UNIV OF GLASGOW.
PA (EQUI-) EQUINE VIROLOGY RES FOUN.
PI Nicolson L, Onions DE;
DR WPI; 92-056872/07.
DR P-PSDB; R20796.
PT Nucleic acid sequence encoding EHV-4 gH or gC protein - used to
PT produce a vaccine for protection of horses against EHV-4
PT infection
PS Claim 1; Page 23; 29pp; English.
CC Equine dermal cells (NBL-6) were infected with EHV-4 strain 1942
CC viral DNA, purified and a BamHI library constructed in pUC9.
CC Calcium shocked E. coli DH7 cells were transformed with the
CC recombinant plasmids. Additional clones were derived from a
CC restriction digest of pUC9 contg. the BamHI G fragment. The
CC nucleotide sequence of a region of BamHI G fragment spanning the gC
CC gene was determined. by analysis of overlapping sequences (SEQ ID no
CC 2). Vaccines can be prepd. using this sequence, and they may be used
CC to protect horses against EHV-4 infection, inducing a higher level
CC of immunity and less side-effects than other live virus vaccines.
CC See also Q20994.
SQ Sequence 1560 BP;   438 A;   409 C;   334 G;   379 T;

alignment_scores:
        Quality: 34.00   Length: 10
        Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x Q20995   ..

Align seg 1/1 to: Q20995 from: 1 to: 1560

     1 TyrArgLeuIleArgLeuAsnGluArg 10
       |||||||| |||::||||||::|||
   409 TACAGACTAGAAATTCACCTAAACCAGCGC 438

seq_name: N_Geneseq_36:V82077

seq_documentation_block:
ID V82077 standard; DNA; 2132 BP.
AC V82077;
DT 25-MAR-1999 (first entry)
DE DNA encoding a gidAl protein.
KW GidAl; bacterial infection; meningitis; Helicobacter pylori infection;
KW cancer; ulcer; gastritis; antibacterial; in-dwelling device;
KW wound treatment; bacterial adhesion; matrix protein; ds.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT CDS   109..1986
FT   /*tag= a
FT   /product= gidAl
PN EP-889129-A2.
PD 07-JAN-1999.
PF 30-JUN-1998; 305180.
PR 01-JUL-1997; US-052758.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Burnham M, Kallender H, Lenox AL, Palmer LM;
DR WPI; 99-062660/06.

DR P-PSDB; W89445.
PT New isolated gidAl polypeptide from Staphylococcus aureus - used to
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT H. pylori and associated cancers, ulcers and gastritis
PS Claim 2; Page 3-4; 43pp; English.
CC The present sequence encodes a gidAl protein of Staphylococcus
CC aureus. GidAl proteins, nucleic acids and agonists are used to
CC treat conditions requiring increased activity or expression of gidAl,
CC while conditions (particularly bacterial infections) requiring
CC inhibition of gidAl are treated by administering an antagonist,
CC inhibitory nucleic acid or competitive polypeptide. The products are
CC used to treat S. pneumoniae infection, particularly meningitis and
CC also Helicobacter pylori infections e.g. related cancers, ulcers and
CC gastritis. These antibacterial agents may also be used to treat
CC in-dwelling devices to prevent infection or generally as wound
CC treatments to prevent adhesion of bacteria to matrix proteins.
SQ Sequence 2132 BP;   791 A;   311 C;   430 G;   600 T;

alignment_scores:
        Quality: 34.00   Length: 10
        Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-11 x V82077   ..

Align seg 1/1 to: V82077 from: 1 to: 2132

     1 TyrArgLeuIleArgLeuAsnGluArg 10
       ||||||||||||:::::: |||||:::
  2070 TATCGTTTACTTGTTGAATGGAATGAAAAG 2099

OM of: US-08-653-294-11 to: EST:*   out_format : pfs
Date: Feb 8, 2000 4:02 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1/USPTO_spool/US08653294/runat_04022000_160700_15770/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-11
Query length: US-08-653-294-11
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 8553.360000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est40:AV242923	+	40.00	147.78	18.24	246	! AV242923 AV242923 RIKEN full-le
gb_gss4:AQ735952	+	39.00	137.15	71.25	578	! AQ735952 HS_2261_A1_F09_T7C CIT
gb_gss1:CNS002P7	+	39.00	132.96	121.95	983	! AI097957 Drosophila melanogaste
gb_gss10:AQ157379	+	38.00	132.13	135.72	668	! AQ157379 nbxb0009l8r CUGI Rice
gb_gss11:AQ271001	-	38.00	129.38	193.01	946	! AQ271001 nbxb0015al1f CUGI Rice
gb_gss1:CNS00G7E	+	38.00	128.18	225.05	1101	! AL071730 Drosophila melanogast
gb_gss1:CNS00PG6	+	37.00	132.31	132.55	399	! AL084468 Arabidopsis thaliana g
gb_gss8:AQ064329	+	37.00	132.23	133.90	403	! AQ064329 HS_2200_A1_MF_E07 CIT
gb_est22:AI008780	+	37.00	132.19	134.17	405	! AI008780 ESTO3231 Normalized r
gb_est14:AQ583741	-	37.00	129.75	184.11	552	! AQ583741 RPCI-11-434G5.TV RPCI
gb_est35:AI820472	-	36.00	131.99	138.09	254	! AI820472 ME000422.FOR Egg stage
gb_est10:AA185851	+	36.00	130.57	165.63	304	! AA185851 SW3lCAl954SK Brugia ma
gb_est17:AA604578	+	36.00	130.05	177.21	325	! AA604578 no76b01.s1 NCI_CGAP_AP
gb_est32:AI758190	-	36.00	129.76	183.84	337	! AI758190 ty70e05.x1 NCI_CGAP_Ki
gb_est36:AV188192	-	36.00	129.24	196.54	360	! AV188192 AV188192 Yuji Kohara u
gb_gss9:AQ128402	+	36.00	128.83	207.04	379	! AQ128402 HS_3085_A1_E07_MR CIT
gb_est26:AI350192	+	36.00	128.01	230.28	421	! AI350192 qo28h01.x1 NCI_CGAP_Lu
gb_gss8:AQ675126	-	36.00	127.78	236.92	433	! AQ675126 HS_2162_A1_D10_T7C CIT
gb_gss4:AQ054233	-	36.00	127.66	238.03	435	! AQ054233 CIT-HSP-2342I20.TR CIT
gb_est31:F32239	-	36.00	127.66	240.80	440	! F32239 HSPD24851 HM3 Homo sapie
gb_gss3:B57891	+	36.00	127.50	245.78	449	! B57891 CIT-HSP-2010015.TR CIT-H
gb_gss6:AQ886481	+	36.00	127.50	245.78	449	! AQ886481 HS_543_B2_A12_SP6E RP
gb_est16:AA557396	-	36.00	127.39	249.11	455	! AA557396 nl81h06.s1 NCI_CGAP_Br
gb_gss9:AQ146935	-	36.00	127.39	249.11	455	! AQ146935 HS_2248_A2_B10_MR CIT
gb_gss10:AQ215203	+	36.00	127.39	249.11	455	! AQ215203 HS_2010_A1_H08_MR CIT
gb_est10:AQ216648	+	36.00	127.24	254.10	464	! AQ216648 HS_2190_A1_B03_T7 CIT
gb_gss4:AQ685122	+	36.00	126.60	275.72	503	! AQ685122 HS_1113_B2_E07_T7 CIT
gb_est33:AI772130	+	36.00	126.43	281.83	514	! AI772130 HS_2160_A1_F05_T7C CIT
gb_est33:AI779338	+	36.00	126.31	286.27	522	! AI772130 EST253230 tomato resis
gb_est24:AU031017	+	36.00	125.41	321.26	585	! AI779338 EST260217 tomato susce
gb_est30:AI658427	+	36.00	124.68	352.96	642	! AU031017 Rice cDNA fro
gb_est30:AI658444	+	36.00	124.40	365.76	665	! AI658427 AEMTBM58 Aedes aegypti
gb_gss1:CNS008I	+	36.00	124.27	371.88	676	! AI658444 AEMTBN57 Aedes aegypti
gb_est21:AA911731	-	36.00	122.44	470.02	852	! AL051558 Drosophila melanogaste
gb_est23:AI128565	-	35.00	133.90	108.16	122	! AA911731 o15e06.s1 NCI_CGAP_GC
gb_gss1:CNS00S22	-	35.00	132.43	130.62	147	! AI128565 qa61b01.s1 Soares_feta
gb_est1:F02347	-	35.00	130.24	172.97	194	! AL087848 Arabidopsis thaliana g
gb_est8:AA026266	-	35.00	129.72	184.70	207	! F02347 HSCOWG062 normalized inf
gb_est42:AV294974	+	35.00	129.07	200.97	225	! AA026266 ze97f02.s1 Soares_feta
gb_est35:AI866895	+	35.00	128.96	203.68	228	! AV294974 AV294974 RIKEN full-le
gb_est18:AA729926	-	35.00	128.52	215.44	241	! AI866895 wn12e12.x1 NCI_CGAP_Ga
gb_est18:AA729926	-	35.00	127.43	248.04	277	! AA729926 nx4f01.s1 NCI_CGAP_GO
gb_est20:AA864515	-	35.00	127.20	255.29	285	! AA864515 oh55b03.s1 NCI_CGAP
gb_est24:AI233666	+	35.00	127.15	257.10	287	! AI233666 EST233354 Normalize
gb_est25:AI246027	-	35.00	126.98	262.54	293	! AI246027 qk44g11.x1 NCI_CGAP

seq_name: gb_est40:AV242923

seq_documentation_block:
LOCUS   AV242923   246 bp   mRNA   EST   04-NOV-1999
DEFINITION   AV242923 RIKEN full-length enriched, 0 day neonate head Mus
             musculus cDNA clone 4831414O15 3' similar to NM_004576 Homo sapiens
             protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52),
             beta isoform (PPP2R2B) mRNA, mRNA, mRNA sequence.
ACCESSION   AV242923
VERSION   AV242923.1 GI:6230332
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 246)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
             Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
             Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
             Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
             Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
             Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
             Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
             Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
             Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
             Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT   On May 18, 1998 this sequence version replaced gi:3137751.
             Contact: Yoshihide Hayashizaki
             Genome Exploration Research Group, Life Science Tsukuba Center,
             Genome Science Laboratory
             The Institute of Physical and Chemical Research (RIKEN), Genomic
             Sciences Center
             3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
             Tel: +81-298-36-9013
             Fax: +81-298-36-9098
             Email: genome-res@rtc.riken.go.jp,
             URL:http://genome.rtc.riken.go.jp/
             Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
             Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
             Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
             polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
             Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
             Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
             system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
             19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
             further details.
FEATURES   Location/Qualifiers
     source   1..246
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4831414O15"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGGATCCAAGACTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAGATTCGAGTAATTAAATTAATCCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I."

```
BASE COUNT          71 a      39 c      37 g      99 t
ORIGIN
```

alignment_scores:
```
     Quality:    40.00           Length:      9
       Ratio:     4.444            Gaps:       0
Percent Similarity: 100.000   Percent Identity: 88.889
```

alignment_block:
US-08-653-294-11 x AV242923
Align seg 1/1 to: AV242923 from: 1 to: 246

```
  1 TyrArgLeuLeuIleArgLeuAsnGlu 9
    ||||:::|||||||||||||||
140 TATAGGTTTCTCATCCGATTAAATGAG 166
```

seq_name: gb_gss4:AQ735952

seq_documentation_block:
```
LOCUS       AQ735952         578 bp    DNA           GSS     15-JUL-1999
DEFINITION  HS_2261_A1_F09_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2261 Col=17 Row=K, genomic survey
            sequence.
ACCESSION   AQ735952
VERSION     AQ735952.1  GI:5507504
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
   MEDLINE   99380589
  COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2261  row: K  column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1..578
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2261 Col=17 Row=K"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
```

/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC clones in
E-Coli DH10B"

```
BASE COUNT         207 a      90 g      97 c     180 t      4 others
ORIGIN
```

alignment_scores:
```
     Quality:    39.00            Length:     10
       Ratio:     3.900            Gaps:       0
Percent Similarity: 100.000   Percent Identity: 60.000
```

alignment_block:
US-08-653-294-11 x AQ735952
Align seg 1/1 to: AQ735952 from: 1 to: 578

```
  1 TyrArgLeuLeuIleArgLeuAsnGluArg 10
    |||||||||:::::|||:::||:::||
 94 TACAGATTAATCATGAGGATTAAATGACAGA 123
```

seq_name: gb_gss1:CNS002P7

seq_documentation_block:
```
LOCUS       CNS002P7          983 bp    DNA           GSS     26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN02A09 of DrosBAC library from Drosophila melanogaster (fruit
            fly) genomic survey sequence.
ACCESSION   AL097957
VERSION     AL097957.1  GI:5609568
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
  COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..983
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN02A09"
                     /note="end : SP6"
```

```
BASE COUNT         313 a     152 c     178 g     276 t     64 others
ORIGIN
```

alignment_scores:
```
     Quality:    39.00            Length:     10
       Ratio:     3.900            Gaps:       0
Percent Similarity: 100.000   Percent Identity: 70.000
```

alignment_block:
US-08-653-294-11 x CNS002P7/rev
Align seg 1/1 to reverse of: CNS002P7 from: 1 to: 983

1 TyrArgLeuIleLeuArgLeuAsnGluArg 10

```
Align seg 1/1 to: AQ157379 from: 1 to: 668

  1 TyrArgLeuIleArgLeuAsnGluArg 10
    |||||||:::::||||||||||::|||
526 TATCGTCTTATGTTTAGACTTAATAGCCGT 555

seq_name: gb_gss11:AQ271001

seq_documentation_block:
LOCUS       AQ271001      946 bp    DNA             GSS       03-NOV-1998
DEFINITION  nbxb0015A11f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0015A11f, genomic survey sequence.
ACCESSION   AQ271001
VERSION     AQ271001.1   GI:3824316
KEYWORDS    GSS.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 946)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence stop: 152.
FEATURES             Location/Qualifiers
     source          1..946
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone_lib="nbxb0015A11f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep coverage allows the isolation a particular
                     sequence with a probability of 99.9 %. Two high density
                     filters, each containing 18,432 clones (doubly spotted),
                     represent the whole library for colony screening."
BASE COUNT  260 a    202 c    181 g    303 t
ORIGIN

alignment_scores:
    Quality:     38.00             Length:        10
    Ratio:        4.222            Gaps:           0
Percent Similarity:  90.000     Percent Identity:  80.000
```

```
    |||::::||||||||||:::||::::|||
626 TATAAACTCTTGATCAGGATCAATAGCCGA 597

seq_name: gb_gss10:AQ157379

seq_documentation_block:
LOCUS       AQ157379      668 bp    DNA             GSS       12-SEP-1998
DEFINITION  nbxb0009I18r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0009I18r, genomic survey sequence.
ACCESSION   AQ157379
VERSION     AQ157379.1   GI:3592495
KEYWORDS    GSS.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 668)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3554404.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1..668
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone_lib="nbxb0009I18r"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep coverage allows the isolation a particular
                     sequence with a probability of 99.9 %. Two high density
                     filters, each containing 18,432 clones (doubly spotted),
                     represent the whole library for colony screening."
BASE COUNT  157 a    204 c    141 g    166 t
ORIGIN

alignment_scores:
    Quality:     38.00             Length:        10
    Ratio:        3.800            Gaps:           0
Percent Similarity:  100.000    Percent Identity:  70.000

alignment_block:
US-08-653-294-11 x AQ157379
```

alignment_block:
US-08-653-294-11 x AQ271001/rev

Align seg 1/1 to reverse of: AQ271001   from: 1 to: 946

```
  1 TyrArgLeuIleArgLeuAsnGluArg 10
    |||::||||||||  ||||||||||||
732 TATCAGCTATTAATAGATCTAAATGACGGT 703
```

seq_name: gb_gss1:CNS00G7E

seq_documentation_block:
```
LOCUS       CNS00G7E     1101 bp    DNA         GSS    03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR32013 of RPCI-98 library from Drosophila melanogaster (fruit
            fly),genomic survey sequence.
ACCESSION   AL071730
VERSION     AL071730.1    GI:4951570
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR32013"
                     /note="end : T7"
BASE COUNT      305 a    247 c    220 g    294 t
ORIGIN
```

alignment_scores:
```
      Quality:     38.00              Length:     10
      Ratio:        3.800             Gaps:        0
Percent Similarity: 100.000     Percent Identity: 70.000
```

alignment_block:
US-08-653-294-11 x CNS00G7E/rev

Align seg 1/1 to reverse of: CNS00G7E   from: 1 to: 1101

```
  1 TyrArgLeuIleArgLeuAsnGluArg 10
    |||::||||||:::||||||||||
925 TACCACCTTCTTTTGCAGCTGAACGAGAGG 896
```

seq_name: gb_gss1:CNS00PG6

seq_documentation_block:

```
LOCUS       CNS00PG6      399 bp    DNA         GSS    28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence T7 end of BAC F801 of
            IGF library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL08468
VERSION     AL08468.1    GI:5285608
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 399)
AUTHORS     Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 399)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1..399
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="IGF"
                     /clone="F801"
                     /note="end : T7"
BASE COUNT      105 a    57 c    98 g    139 t
ORIGIN
```

alignment_scores:
```
      Quality:     37.00              Length:      8
      Ratio:        4.625             Gaps:        0
Percent Similarity: 100.000     Percent Identity: 87.500
```

alignment_block:
US-08-653-294-11 x CNS00PG6

Align seg 1/1 to: CNS00PG6   from: 1 to: 399

```
  1 TyrArgLeuIleArgLeuAsn 8
    |||||:::||||||||||||
207 TACCGATTGATAATTAGGTTGAAT 230
```

seq_name: gb_gss8:AQ064329

seq_documentation_block:
```
LOCUS       AQ064329      403 bp    DNA         GSS    04-AUG-1998
DEFINITION  HS_2200_A1_MF_E07 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2200 Col=13 Row=I, genomic survey
            sequence.
ACCESSION   AQ064329
VERSION     AQ064329.1    GI:3378867
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 403)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
```

```
                    University of Washington
                    401 Queen Anne Avenue North, Seattle, WA 98109, USA
                    Tel: (206) 616-3618
                    Fax: (206) 616-3887
                    Email: jwallace@u.washington.edu
                    Sequence Tagged Connector
                    Plate: 2200 row: I column: 13
                    Class: BAC ends
                    High quality sequence stop: 403.
FEATURES            Location/Qualifiers
     source         1..403
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="plate-2200 Col=13 Row=I"
                    /clone_lib="CIT Approved Human Genomic Sperm Library D"
                    /sex="male"
                    /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
BASE COUNT    100 a    88 c    82 g   132 t    1 others
ORIGIN

alignment_scores:
     Quality:      37.00              Length:      10
     Ratio:         3.700             Gaps:         0
Percent Similarity: 100.000     Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AQ064329/rev     ..

Align seg 1/1 to reverse of: AQ064329   from: 1  to: 403

    1 TyrArgLeuIleLeuArgLeuAsnGluArg 10
      ||||||||||||||||||||||||::||||||
  395 TATCGCCTTTAATCCAAATTGTGAAAGA 366

seq_name: gb_est22:AI008780

seq_documentation_block:
LOCUS       AI008780        405 bp   mRNA              EST   25-JAN-1999
DEFINITION  EST203231 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMBD17 3' end, mRNA sequence.
ACCESSION   AI008780
VERSION     AI008780.1   GI:3222612
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 405)
AUTHORS     Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE       Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL     Unpublished (1998)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044445.
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Contact: Lee, NH
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES            Location/Qualifiers
     source         1..405
                    /organism="Rattus sp."
                    /db_xref="ATCC (inhost):2016965"
                    /db_xref="taxon:10118"
                    /clone="REMBD17"
                    /clone_lib="Normalized rat embryo, Bento Soares"
                    /dev_stage="embryo 8, 12, 18 dpc"
                    /note="vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
```

```
BASE COUNT    117 a    74 c    63 g   151 t
ORIGIN

alignment_scores:
     Quality:      37.00              Length:       9
     Ratio:         4.111             Gaps:         0
Percent Similarity: 100.000     Percent Identity: 66.667

alignment_block:
US-08-653-294-11 x AI008780     ..

Align seg 1/1  to: AI008780   from: 1  to: 405

    1 TyrArgLeuIleLeuArgLeuAsnGlu 9
      ||||||||||||||||||||||||||||
  141 TACAAAATACTTTGAGGTAAATGAA 167

seq_name: gb_gss14:AQ583741

seq_documentation_block:
LOCUS       AQ583741        552 bp   DNA               GSS   07-JUN-1999
DEFINITION  RPCI-11-434G5.TV RPCI-11 Homo sapiens genomic clone RPCI-11-434G5,
            genomic survey sequence.
ACCESSION   AQ583741
VERSION     AQ583741.1   GI:5010851
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
TITLE       Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet.cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1..552
                    /organism="Homo sapiens"
                    /db_xref="GDB:7666420"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-434G5"
                    /clone_lib="RPCI-11"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
BASE COUNT    144 a   117 c    70 g   220 t    1 others
ORIGIN

alignment_scores:
     Quality:      37.00              Length:      10
     Ratio:         3.700             Gaps:         0
Percent Similarity: 100.000     Percent Identity: 60.000

alignment_block:
```

US-08-653-294-11 x AQ583741/rev

Align seg 1/1 to reverse of: AQ583741 from: 1 to: 552

```
    1 TyrArgLeuLeuIleArgLeuAsnGluArg 10
      ||||||||||::::::||:::||:::|||
  350 TACAGGCTAATTTTAGAATTAACAAGAGA 321
```

seq_name: gb_est35:AI820472

```
seq_documentation_block:
LOCUS       AI820472    254 bp   mRNA         EST      09-JUL-1999
DEFINITION  ME000422.FOR Egg stage cDNA expression library in Lambda ZAPII
            Schistosoma mansoni cDNA 5', mRNA sequence.
ACCESSION   AI820472
VERSION     AI820472.1  GI:5439551
KEYWORDS    EST.
SOURCE      Schistosoma mansoni.
ORGANISM    Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1 (bases 1 to 254)
  AUTHORS   Oliveira,G.C. and Baba,J.
  TITLE     Cataloguing Schistosoma mansoni genes with expressed sequence tags
  JOURNAL   Unpublished (1998)
COMMENT     On Feb 18, 1999 this sequence version replaced gi:4296741.
            Contact: Oliveira, Guilherme
            Lab. Parasitologia Cel. e Mol.
            Centro de Pesquisas Rene Rachou - FIOCRUZ
            Av. Augusto de Lima 1715 Barro Preto, Belo Horizonte, MG, CEP
            30190, Brazil
            Tel: 55 31 2953566
            Fax: 55 31 2952115
            Email: oliveira@netra.cpqrr.fiocruz.br
            Insert Length: 254  Std Error: 0.00
            Seq primer: M13 Reverse Universal Sequencing primer.
FEATURES             Location/Qualifiers
     source          1..254
                     /organism="Schistosoma mansoni"
                     /strain="LE"
                     /db_xref="taxon:6183"
                     /clone_lib="Egg stage cDNA expression library in Lambda
                     ZAPII"
                     /sex="Mixed"
                     /dev_stage="Egg"
                     /note="Vector: Uni-Zap XR vector, Stratagene (pBluescript
                     SK); Site_1: EcoRI; Site_2: XhoI; mRNA was extracted from
                     eggs and the library was constructed and excised
                     according to the manufacturer's instructions."
BASE COUNT     67 a     16 c     69 g    102 t
ORIGIN
```

alignment_scores:
```
        Quality:       36.00          Length:        9
        Ratio:          4.000          Gaps:          0
Percent Similarity: 100.000   Percent Identity: 88.889
```

alignment_block:
US-08-653-294-11 x AI820472/rev

Align seg 1/1 to reverse of: AI820472 from: 1 to: 254

```
    2 ArgLeuLeuIleArgLeuAsnGluArg 10
      ||||||||||::::||||||||||
  161 AGGTTACTTATCCGACTCACCGAACGA 135
```

seq_name: gb_est10:AA185851

```
seq_documentation_block:
LOCUS       AA185851   304 bp   mRNA        EST     08-JAN-1997
DEFINITION  SW3ICAI9545K Brugia malayi infective larva cDNA (SAW94WL-BmL3)
            Brugia malayi cDNA clone SW3ICAI954 5', mRNA sequence.
ACCESSION   AA185851
VERSION     AA185851.1  GI:1769739
KEYWORDS    EST.
SOURCE      Brugia malayi.
ORGANISM    Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1 (bases 1 to 304)
  AUTHORS   Williams,S.A.
  TITLE     Genes expressed in infective third stage larvae of Brugia malayi
  JOURNAL   Unpublished (1995)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1325247.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1..304
                     /organism="Brugia malayi"
                     /strain="TRS Labs"
                     /db_xref="taxon:6279"
                     /clone="SW3ICAI954"
                     /clone_lib="Brugia malayi infective larva cDNA
                     (SAW94WL-BmL3)"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: lambda UniZap XR: Site_1: EcoR I: Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from third stage infective larvae of
                     Brugia malayi isolated from mosquitoes 10 days after
                     infection and converted to double stranded cDNA using
                     reverse transcriptase and oligo(dT) followed by RNase H
                     and DNApol I. The library had 1.6 x 10E6 independent
                     recombinants and average insert size was 900 base pairs.
                     The library was constructed by Wenhong Lu. The library is
                     available from Dr. S.A. Williams, email genome@smith.edu."
BASE COUNT     79 a     78 c     70 g     73 t      4 others
ORIGIN
```

alignment_scores:
```
        Quality:       36.00          Length:       10
        Ratio:          3.600          Gaps:          0
Percent Similarity: 100.000   Percent Identity: 60.000
```

alignment_block:
US-08-653-294-11 x AA185851

Align seg 1/1 to: AA185851 from: 1 to: 304

```
    1 TyrArgLeuLeuIleArgLeuAsnGluArg 10
      |||||||||:||::::||||:::|||  |||
  203 TATCGTCTGATTGTCCGTTTGAGCAACCGT 232
```

seq_name: gb_est17:AA604578

```
seq_documentation_block:
LOCUS       AA604578   325 bp   mRNA        EST      08-OCT-1997
DEFINITION  no76b01.s1 NCI_CGAP_AA1 Homo sapiens cDNA clone IMAGE:1112713 3',
            mRNA sequence.
ACCESSION   AA604578
VERSION     AA604578.1  GI:2445442
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 325)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1405116.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
            Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image.html

            Insert Length: 682    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 217.
FEATURES             Location/Qualifiers
     source          1..325
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1112713"
                     /clone_lib="NCI_CGAP_AA1"
                     /tissue_type="adrenal adenoma"
                     /note="Organ: adrenal gland; Vector: Bluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
                     Primer: Oligo dT. Two pooled bulk adrenal adenomas. 5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTTT 3' Average insert
                     size: 1.6 kb."
BASE COUNT      113 a     46 c     57 g     109 t
ORIGIN

alignment_scores:
        Quality:     36.00              Length:        10
        Ratio:        3.600             Gaps:           0
    Percent Similarity: 100.000     Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AA604578    ..

Align seg 1/1   to: AA604578   from: 1   to: 325

    1 TyrArgLeuIleArgLeuAsnGluArg 10
      |||:||||||||||||||||||||:::::
  172 TACAATCTACTCATAAGATTAAACCACAAA 201

seq_name: gb_est32:AI758190

seq_documentation_block:
LOCUS       AI758190        337 bp    mRNA          EST      23-JUN-1999
DEFINITION  ty70e05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2284448 3',
            mRNA sequence.
ACCESSION   AI758190
VERSION     AI758190.1   GI:5151913
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133827.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..337
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2284448"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Site_1: Not I; Site_2: Eco RI;
                     a modified polylinker; Vector: pT7T3D-Pac (Pharmacia) with
                     plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIds 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo."
BASE COUNT      131 a     50 c     55 g     101 t
ORIGIN

alignment_scores:
        Quality:     36.00              Length:        10
        Ratio:        3.600             Gaps:           0
    Percent Similarity: 100.000     Percent Identity: 70.000

alignment_block:
US-08-653-294-11 x AI758190    ..

Align seg 1/1   to: AI758190   from: 1   to: 337

    1 TyrArgLeuIleArgLeuAsnGluArg 10
      |||:||||||||||||||||||||:::::
  242 TACAATCTACTCATAAGATTAAACCACAAA 271

seq_name: gb_est36:AV188192

seq_documentation_block:
LOCUS       AV188192        360 bp    mRNA          EST      22-JUL-1999
DEFINITION  AV188192 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk520g12 5', mRNA
            sequence.
ACCESSION   AV188192
VERSION     AV188192.1   GI:5570175
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
TITLE       Expressed genes in C.elegans
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137896.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
```

```
           Fax: 81-559-81-6855
           Email: ykohara@lab.nig.ac.jp.
FEATURES          Location/Qualifiers
     source       1..360
                  /organism="Caenorhabditis elegans"
                  /strain="N2"
                  /db_xref="taxon:6239"
                  /clone="yk520g12"
                  /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                  hermaphrodite embryo"
                  /sex="hermaphrodite"                .
                  /dev_stage="embryo"
BASE COUNT        94 a ,  92 c     37 g     137 t
ORIGIN

alignment_scores:
       Quality:    36.00         Length:         8
       Ratio:       4.500         Gaps:          0
Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
US-08-653-294-11 x AV188192/rev     ..

Align seg 1/1  to reverse of: AV188192  from: 1   to: 360


   1 TyrArgLeuIleArgLeuAsn 8
     ||:::|||||||||||||||
 186 TATAAACTTTGATTAGATTAAAT 163
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 8, 2000, 01:29:38 ; Search time 122.56 seconds
                                         (without alignments)
                                         1.933 Million cell updates/sec

Title:           US-08-653-294-13
Perfect score:   49
Sequence:        1 YRLAIRLDER 10

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:    188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :    A_Geneseq_36:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	10	1	W47270	Immunomodulatory p
2	44	89.8	10	1	W47266	Immunomodulatory p
3	44	89.8	10	1	W47272	Immunomodulatory p
4	44	89.8	20	1	R92909	HLA-B2702 CTL modu
5	44	89.8	20	1	R92911	HLA-B2702 CTL modu
6	44	89.8	20	1	R92907	HLA-B2702 CTL modu
7	44	89.8	20	1	R95428	HLA-B2702 84-75-84
8	44	89.8	20	1	W33778	Immunomodulating d
9	44	89.8	20	1	W33779	Immunomodulating d
10	44	89.8	20	1	W33792	Peptide B2702.84-7
11	39	79.6	10	1	W47268	Immunomodulatory p
12	39	79.6	20	1	R92910	HLA-B2702 CTL modu
13	39	79.6	20	1	R92908	HLA-B2702 CTL modu
14	39	79.6	20	1	R95430	HLA-B2702 84-75T/7
15	39	79.6	20	1	W33791	Peptide B2702.84-7
16	39	79.6	20	1	W33793	Peptide B2702.84-7
17	32	65.3	318	1	R88105	Bovine p32 11-cis-
18	32	65.3	318	1	W18343	Bovine p32 11-cis-
19	32	65.3	318	1	Y01994	A p32 protein whic
20	32	65.3	546	1	W26355	KOD-1 heat shock p
21	32	65.3	548	1	W67797	Thermococcus sp. K
22	31	63.3	485	1	R20796	EHV-4 gC. Nucleic
23	30	61.2	3398	1	R44430	eryA region polype
24	30	61.2	3457	1	R62504	Large polyprotein
25	30	61.2	3588	1	R34712	Bacillus subtilis
26	29	59.2	6	1	W47263	Immunomodulatory p
27	29	59.2	6	1	W33781	Peptide #2 used in
28	29	59.2	12	1	R95429	HLA-B2702 84-79-84
29	29	59.2	12	1	W33798	Peptide B2702.84-7
30	29	59.2	12	1	W33799	Immunomodulating d
31	29	59.2	53	1	W19361	Beta 7 integrin S3
32	29	59.2	158	1	W23586	Mabinlin MBLI from
33	29	59.2	158	1	W23588	Mabinlin MBLIII fr
34	29	59.2	219	1	W82592	Mouse Rit ras-like
35	29	59.2	376	1	R21416	Carbonic anhydrase
36	29	59.2	377	1	R21417	Chlamydomonas carb
37	29	59.2	381	1	W98786	H. pylori GHPO 121
38	29	59.2	537	1	R25534	Toxoplasma gondii
39	29	59.2	546	1	W67798	Thermococcus sp. K
40	29	59.2	663	1	W09055	Nicotiana plumbagi
41	29	59.2	711	1	W55103	Streptococcus pneu
42	29	59.2	738	1	R69849	Ethylene response
43	29	59.2	738	1	R69852	Ethylene response
44	29	59.2	738	1	R69853	Ethylene response
45	29	59.2	738	1	W73121	A. thaliana ethyle

ALIGNMENTS

```
RESULT   1
W47270
ID   W47270 standard; peptide; 10 AA.
AC   W47270;
DT   22-MAY-1998  (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   Misc_difference 1. .10
FT                   /note= "at least one of the amino acids is the
FT                   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence  10 AA;

Query Match            100.0%;    Score 49;  DB 1;  Length 10;
Best Local Similarity  100.0%;    Pred. No. 0.00014;
Matches   10;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 YRLAIRLDER 10
            ||||||||||
Db      1 YRLAIRLDER 10


RESULT   2
W47266
ID   W47266 standard; peptide; 10 AA.
AC   W47266;
DT   22-MAY-1998  (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   Misc_difference 1. .10
```

FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence 10 AA;

Query Match   89.8%;   Score 44; DB 1; Length 10;
Best Local Similarity 90.0%;   Pred. No. 0.0016;
Matches   9; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||||||:||
Db   1 YRLAIRLNER 10

RESULT 3
W47272
ID   W47272 standard; peptide; 10 AA.
AC   W47272;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence 10 AA;

Query Match   89.8%;   Score 44; DB 1; Length 10;
Best Local Similarity 90.0%;   Pred. No. 0.0016;
Matches   9; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||| |||||
Db   1 YRLLIRLDER 10

RESULT 4
R92909
ID   R92909 standard; peptide; 20 AA.
AC   R92909;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Exmple 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   89.8%;   Score 44;   DB 1; Length 20;
Best Local Similarity 90.0%;   Pred. No. 0.0035;
Matches   9; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||||||:||
Db   1 YRLAIRLNER 10

RESULT 5
R92911
ID   R92911 standard; peptide; 20 AA.
AC   R92911;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/84-75).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is a dimer of residues 75-84 of the alpha-1 domain of the class I MHC

CC HLA-B2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 89.8%; Score 44; DB 1; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.0035;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRLAIRLDER 10
      ||||||:||
Db 1 YRLAIRLNER 10

RESULT 6
R92907
ID R92907 standard; peptide; 20 AA.
AC R92907;
DT 16-MAY-1996 (first entry)
DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW Cytotoxic T lymphocyte: CTL: major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B2702.
OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI: 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of the class I MHC alpha-1 domain of the class
CC I MHC HLA-B2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 89.8%; Score 44; DB 1; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.0035;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRLAIRLDER 10
      ||||||:||
Db 1 YRLAIRLNER 10

RESULT 7
R95428
ID R95428 standard; peptide; 20 AA.
AC R95428;
DT 12-NOV-1996 (first entry)
DE HLA-B2702 84-75-84 palindrome.
KW HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW T-cell lysate: membrane protein: mammal: heat shock protein: Hsc70: APC;
KW B cell; calcium influx; cytotoxic T lymphocyte: CTL; differentiation;

KW cytolysis: antigen presenting cell.
OS Synthetic.
PN WO9513288-A1.
PD 18-MAY-1995.
PF 10-NOV-1994; U12985.
PR 10-NOV-1993; US-150493.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI: 95-194027/25.
PT Compsns. comprising lymphoid surface membrane proteins - which may
PT inhibit cytolytic activity and differentiation of CTLs.
PS Example: Page 12; 29pp; English.
CC R95413, and R95415-R95431 represent palindromes and fragments of
CC human-leucocyte-associated antigens. This sequence represents the
CC HLA-B2702 84-75-84 palindrome. These sequences can be used to isolate
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC protein associated with T-cell activation in mammalian T-cells, and is
CC also immunologically cross reactive with the heat shock protein Hsc70.
CC p74 is found in a limited number of cell types, but is particularly
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC cell with an amphoteric detergent, and then passed through an affinity
CC column containing a covalently bound HLA-B2702 palindromic peptide.
CC Compositions comprising the extracellular fragment of p74 combined with
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC compounds can be screened for their effect on the cytolytic activity of
CC T-cells, by combining them with the extracellular portion of p74 and
CC determining the amount of binding between the candidate compound and p74.
CC Modulation of CTL activity can be inhibited in a cellular composition
CC containing T-cells and antigen presenting cells (APCs), by adding to the
CC mix the extracellular portion of p74, in an amount sufficient to compete
CC with p74 for the binding of the p74 ligand.
SQ Sequence 20 AA;

Query Match 89.8%; Score 44; DB 1; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.0035;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRLAIRLDER 10
      ||||||:||
Db 1 YRLAIRLNER 10

RESULT 8
W33778
ID W33778 standard; peptide; 20 AA.
AC W33778;
DT 19-JUN-1998 (first entry)
DE Immunomodulating dimer peptide #1.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN WO9744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI: 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Claim 16; Page 35; 41pp; English.
CC This sequence represents a specifically claimed immunomodulating
CC dimer peptide of the invention. A peptide-type compound or variant is
CC claimed which has immunomodulating activity, including the N-terminal
CC acylated and/or C-terminal amidated or esterified forms of up to 60
CC amino acids, where the peptide-type compound comprises the formula: A-B,
CC where A, B =(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76= E or
CC V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a

CC   hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha-1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence   20 AA;

Query Match   89.8%;   Score 44; DB 1; Length 20;
Best Local Similarity   90.0%;   Pred. No. 0.0035;
Matches   9; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||||||:||
Db   1 YRLAIRLNER 10

RESULT   9
W33779
ID   W33779 standard; peptide; 20 AA.
AC   W33779;
DT   19-JUN-1998 (first entry)
DE   Immunomodulating dimer peptide #2.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Homo sapiens.
OS   Synthetic.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI: 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 16; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is
CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B = (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or
CC   V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a
CC   hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha-1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence   20 AA;

Query Match   89.8%;   Score 44; DB 1; Length 20;
Best Local Similarity   90.0%;   Pred. No. 0.0035;
Matches   9; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||||||:||
Db   1 YRLAIRLNER 10

RESULT   10
W33792
ID   W33792 standard; peptide; 20 AA.
AC   W33792;
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Homo sapiens.
OS   Synthetic.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI: 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha-1 domain (positions
CC   79-84) from to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   20 AA;

Query Match   89.8%;   Score 44; DB 1; Length 20;
Best Local Similarity   90.0%;   Pred. No. 0.0035;
Matches   9; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
        ||||||:||
Db   1 YRLAIRLNER 10

RESULT   11
W47268
ID   W47268 standard; peptide; 10 AA.
AC   W47268;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.

PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence 10 AA.

Query Match   79.6%; Score 39; DB 1; Length 10;
Best Local Similarity   80.0%; Pred. No. 0.019;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
          ||| ||:|||
Db   1 YRLLIRLNER 10

RESULT 12
R92910
ID   R92910 standard; peptide; 20 AA.
AC   R92910;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   79.6%; Score 39; DB 1; Length 20;
Best Local Similarity   80.0%; Pred. No. 0.04;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
          ||| ||:|||
Db   1 YRLATRLNER 10

RESULT 13
R92908
ID   R92908 standard; peptide; 20 AA.
AC   R92908;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83086 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   79.6%; Score 39; DB 1; Length 20;
Best Local Similarity   80.0%; Pred. No. 0.04;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
          ||| ||:|||
Db   1 YRLATRLNER 10

RESULT 14
R95430
ID   R95430 standard; peptide; 20 AA.
AC   R95430;
DT   12-NOV-1996 (first entry)
DE   HLA-B2702 84-75T/75-84T palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor; Hsc70;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. These sequences represents the
CC   HLA-B2702 84-75T/75-84T palindrome. These sequences can be used to
CC   isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface
CC   membrane protein associated with T-cell activation in mammalian T-cells,
CC   and is also immunologically cross reactive with the heat shock protein
CC   Hsc70. p74 is found in a limited number of cell types, but is
CC   particularly expressed on B and T cells. p74 can be isolated by lysis of

```
CC    a suitable cell with an amphoteric detergent, and then passed through an
CC    affinity column containing a covalently bound HLA-B2702 palindromic
CC    peptide.  Compositions comprising the extracellular fragment of p74
CC    combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and
CC    inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.
CC    Candidate compounds can be screened for their effect on the cytolytic
CC    activity of T-cells, by combining them with the extracellular portion of
CC    p74 and determining the amount of binding between the candidate compound
CC    and p74.  Modulation of CTL activity can be inhibited in a cellular
CC    composition containing T-cells and antigen presenting cells (APCs), by
CC    adding to the mix the extracellular portion of p74, in an amount
CC    sufficient to compete with p74 for the binding of the p74 ligand.
SQ    Sequence   20 AA;

Query Match             79.6%;  Score 39;  DB 1;  Length 20;
Best Local Similarity   88.9%;  Pred. No. 0.04;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 YRLAIRLDE 9
               ||||||:|
Db           1 YRLAIRLNE 9

RESULT  15
W33791
ID    W33791 standard; peptide; 20 AA.
AC    W33791;
DT    19-JUN-1998  (first entry)
DE    Peptide B2702.84-75T/75-84  tested for immunomodulating activity.
KW    Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW    transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW    rejection.
OS    Synthetic.
OS    Homo sapiens.
PN    WO9744351-A1.
PD    27-NOV-1997.
PF    22-MAY-1997; U08689.
PR    24-MAY-1996; US-653294.
PA    (STRD ) UNIV LELAND STANFORD JUNIOR.
PI    Beulow R, Clayberger C, Krensky AM;
DR    WPI: 98-086530/08.
PT    New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT    alpha-1 domain, used for preventing rejection of transplants or
PT    treating autoimmune diseases
PS    Example 1; Page 19; 41pp; English.
CC    Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC    activity. A peptide-type compound or variant is claimed which has
CC    immunomodulating activity, including the N-terminal acylated and/or
CC    c-terminal amidated or esterified forms of up to 60 amino acids, where
CC    the peptide-type compound comprises the formula; A-B, where A, B =
CC    (R aa76-77L)(aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or V; aa77 -
CC    D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a hydrophobic or
CC    small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC    acid. The sequence in the brackets may optionally be absent or truncated
CC    at any peptide type bond within the brackets. The compounds comprise
CC    amino acid sequences related to a Class I HLA-B alphaI domain (positions
CC    79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC    undesirably attacking cells in a host or in vitro.  They can also be
CC    used in combination with antigenic peptides or proteins of interest to
CC    activate CTLs. They can also inhibit the proliferation of T cells in
CC    response to anti-CD3. The peptide can be used for preventing rejection
CC    of transplants or for treating autoimmune diseases, e.g. diabetes,
CC    rheumatoid arthritis and lupus erythematosis. The products can also be
CC    used for detection and diagnosis.
SQ    Sequence   20 AA;

Query Match             79.6%;  Score 39;  DB 1;  Length 20;
Best Local Similarity   80.0%;  Pred. No. 0.04;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           1 YRLAIRLDER 10
               |||| ||:||
Db           1 YRLATRLNER 10


Search completed: February  8, 2000, 01:29:38
Job time: 1750 sec
```

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 7, 2000, 11:54:25 ; Search time 117.7 Seconds
                                         (without alignments)
                                         4.008 Million cell updates/sec

Title:           US-08-653-294-15
Perfect score:   49
Sequence:        1 YRLLIRLDER 10

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        142080 seqs, 47169319 residues

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :       PIR_62:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Total number of hits satisfying chosen parameters: 142080

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	37	75.5	151	2	C71113	probable frxA prot
2	35	71.4	506	2	S37583	RING finger protei
3	35	71.4	513	2	TVHURF	ret finger protein
4	35	71.4	801	4	TVHURE	transforming protein
5	34	69.4	87	2	D70886	hypothetical prote
6	33	67.3	132	2	S75953	hypothetical prote
7	33	67.3	469	2	A71322	hypothetical prote
8	32	65.3	52	2	T07269	hypothetical prote
9	32	65.3	341	2	D48435	cysteine proteinas
10	32	65.3	342	2	S61978	hypothetical prote
11	32	65.3	348	2	I37271	cylicin II - human
12	32	65.3	554	1	F70548	probable menD prot
13	32	65.3	1400	2	B70963	hypothetical prote
14	31	63.3	30	2	S25666	phosphopyruvate hy
15	31	63.3	157	2	H75054	hypothetical prote
16	31	63.3	221	2	S56263	hypothetical prote
17	31	63.3	237	2	H75202	biotin operon repr
18	31	63.3	395	2	T01392	leucine-rich repea
19	31	63.3	560	2	I50372	ORF2 - chicken
20	31	63.3	588	2	E75060	hydrogenase-4 comp
21	31	63.3	825	2	S54465	YTA12 protein prec
22	31	63.3	1489	2	S60416	DNA helicase YGL15
23	31	63.3	1711	1	A47392	chromodomain-helic
24	30	61.2	85	2	S03746	negative sporulati
25	30	61.2	96	2	S03319	hypothetical prote
26	30	61.2	149	2	H75047	f420-nonreducing h
27	30	61.2	156	2	A70968	hypothetical prote
28	30	61.2	173	2	S27599	hypothetical prote
29	30	61.2	198	2	E69779	transcription regu
30	30	61.2	235	2	A71236	probable biotin--[
31	30	61.2	287	2	I39689	hypothetical prote
32	30	61.2	297	2	T12615	ribosomal protein
33	30	61.2	319	2	JC6117	transcription fact
34	30	61.2	324	2	G69515	transcription regu
35	30	61.2	333	2	T05643	hypothetical prote
36	30	61.2	354	1	VGBE67	glycoprotein D pre
37	30	61.2	367	2	T02529	myb-related protei
38	30	61.2	381	2	D64525	GDP-D-mannose dehy
39	30	61.2	391	2	E72539	hypothetical prote
40	30	61.2	404	2	F64238	hypothetical prote
41	30	61.2	447	2	T16527	hypothetical prote
42	30	61.2	456	2	JC6523	26s proteasom p55
43	30	61.2	513	1	I58311	HMG-box containing
44	30	61.2	540	2	T12704	leucine-rich prote
45	30	61.2	595	2	JQ1684	anthranilate synth

ALIGNMENTS

```
RESULT   1
C71113
probable frxA protein - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C;Accession: C71113
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
    M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
DNA Res. 5, 55-76, 1998
A;Reference number: A71000; MUID:98344137
A;Accession: C71113
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-151 <KAW>
A;Cross-references: GB:AP000003; NID:g3236130; PID:d1030708; PID:g3257082
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;gene: PH0674

Query Match      75.5%;  Best Local Similarity   88.9%;  Score 37;  DB 2;  Length 151;
Matches    8;  Conservative    0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy     2 RLLIRLDER 10
           |||| ||||
Db    96 RLLIELDER 104

RESULT   2
S37583
RING finger protein rfp - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
R;Takahashi, M.
submitted to the EMBL Data Library, October 1993
A;Reference number: S37583
A;Accession: S37583
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-506 <TAK>
A;Cross-references: EMBL:X75343; NID:g406747; PIDN:CAA53092.1; PID:g406748
C;Superfamily: rfp transforming protein; RING finger homology
C;Keywords: zinc
F;5-55/Domain: RING finger homology <RNG>

Query Match      71.4%;  Best Local Similarity   77.8%;  Score 35;  DB 2;  Length 506;
Matches    7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Qy    1 YRLLIRLDE 9
        |||| ||:|
Db  191 YRLLARLEE 199

RESULT 3
TVHURF
ret finger protein - human
N;Alternate names: transforming protein rfp
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 18-Jun-1999
C;Accession: A28101
R;Takahashi, M.; Inaguma, Y.; Hiai, H.; Hirose, F.
Mol. Cell. Biol. 8, 1853-1856, 1988
A;Title: Developmentally regulated expression of a human "finger"-containing gene encode
A;Reference number: A28101; MUID:88246464
A;Accession: A28101
A;Molecule type: mRNA
A;Residues: 1-513 <TAK>
A;Cross-references: DDBJ:J03407; NID:g337371; PIDN:AAA36554.1; PID:g337372
C;Genetics:
A;Gene: GDB:RFP
A;Cross-references: GDB:511359; GDB:1391662
A;Map position: 6p22-6p21.3
C;Superfamily: rfp transforming protein; RING finger homology
C;Keywords: DNA binding; transforming protein; zinc
F;1-315/Product: transforming protein rfp (fragment) #status predicted <RET>
F;12-62/Domain: RING finger homology <RNG>
F;16-127/Domain: metal and nucleic acid binding #status predicted <TMN>

Query Match             71.4%;    Score 35;   DB 1;    Length 513;
Best Local Similarity   77.8%;    Pred. No. 16;
Matches   7;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;


Qy    1 YRLLIRLDE 9
        |||| ||:|
Db  198 YRLLARLEE 206

RESULT 4
TVHURE
transforming protein RFP/protein-tyrosine kinase RET mutant fusion protein - human
N;Alternate names: ret oncogene protein
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) ret
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 10-Sep-1997 #text_change 13-Aug-1999
C;Accession: A27203
R;Takahashi, M.; Cooper, G.M.
Mol. Cell. Biol. 7, 1378-1385, 1987
A;Title: ret transforming gene encodes a fusion protein homologous to tyrosine kinases.
A;Reference number: A27203; MUID:87257826
A;Accession: A27203
A;Molecule type: mRNA
A;Residues: 'QAGA',1-801 <TAK>
A;Cross-references: GB:M16029; NID:g340025
A;Note: codons preceding the probable start codon were translated
C;Comment: The ret oncogene is the chimeric product of a translocation mutation between
C;Genetics:
A;Gene: RFP/RET
C;Keywords: ATP; fusion protein; oncogene; phosphotransferase; transforming protein; tyr
F;1-315/Region: transforming protein rfp
F;316-792/Region: protein-tyrosine kinase ret
F;459-467/Region: protein kinase ATP-binding motif
F;487/Active site: Lys #status predicted

Query Match             71.4%;    Score 35;   DB 4;    Length 801;
Best Local Similarity   77.8%;    Pred. No. 26;
Matches   7;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy    1 YRLLIRLDE 9
```

```
Db  198 YRLLARLEE 206

RESULT 5
D70886
hypothetical protein Rv2866 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C;Accession: D70886
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98259987
A;Accession: D70886
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-87 <COL>
A;Cross-references: GB:AL008883; GB:AL123456; NID:g3261490; PID:e1172958; PID:g261281
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2866

Query Match             69.4%;    Score 34;   DB 2;   Length 87;
Best Local Similarity   66.7%;    Pred. No. 4.2;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   1;   Gaps   0;

Qy    1 YRLLIRLDE 9
        ||| |:|
Db   60 YRLLYRIDD 68

RESULT 6
S75953
hypothetical protein - Synechocystis sp. (strain PCC 6803)
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C;Accession: S75953
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S75953
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-132 <KAN>
A;Cross-references: EMBL:D64006; GB:AB001339; NID:g1001291; PID:d1011451; PID:g100131
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match             67.3%;    Score 33;   DB 2;   Length 132;
Best Local Similarity   66.7%;    Pred. No. 10;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   1;   Gaps   0;

Qy    1 YRLLIRLDE 9
        || |:|||:
Db  120 YRKLRLDD 128

RESULT 7
A71322
hypothetical protein TP0471 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
C;Accession: A71322
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
```

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770
A;Accession: A71322
A;Molecule type: DNA
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Residues: 1-469 <COL>
A;Cross-references: GB:AE001223; GB:AE000520; NID:g3322745; PID:g3322760
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0471

Query Match   67.3%;   Score 33;   DB 2;   Length 469;
Best Local Similarity 87.5%;   Pred. No. 38;
Matches   7;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLD 8
          | |||||||
Db   140 YTLLIRLD 147

RESULT 8
T07269
hypothetical protein 52b - Chlorella vulgaris chloroplast
C;Species: chloroplast Chlorella vulgaris
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 22-Jun-1999
C;Accession: T07269
R;Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakas
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
A;Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlo
A;Reference number: Z15985; MUID:97303241
A;Accession: T07269
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-52 <WAK>
A;Cross-references: EMBL:AB001684; NID:d1110444; PID:d1021495
C;Genetics:
A;Genome: chloroplast
C;Keywords: chloroplast

Query Match   65.3%;   Score 32;   DB 2;   Length 52;
Best Local Similarity 60.0%;   Pred. No. 6.4;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLD 10
          : ||: |||
Db   17 FLLLVELDER 26

RESULT 9
D48435
cysteine proteinase AC-3 - nematode (Haemonchus contortus)
C;Species: Haemonchus contortus
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C;Accession: D48435
R;Pratt, D.; Armes, L.G.; Hageman, R.; Reynolds, V.; Boisvenue, R.J.; Cox, G.N.
Mol. Biochem. Parasitol. 51, 209-218, 1992
A;Title: Cloning and sequence comparisons of four distinct cysteine proteases expressed
A;Reference number: A48435; MUID:92244291
A;Accession: D48435
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-341 <PRA>
A;Cross-references: GB:M80388; NID:g159178; PIDN:AAA29178.1; PID:g159179
A;Note: sequence extracted from NCBI backbone (NCBIN:98512, NCBIP:98520)
C;Superfamily: papain

Query Match   65.3%;   Score 32;   DB 2;   Length 341;

Best Local Similarity 60.0%;   Pred. No. 44;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLDER 10
          | |:|:|:|
Db   221 YRKLYRMDKR 230

RESULT 10
S61978
hypothetical protein YPL088w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein LPG20w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 29-Sep-1999
A;Accession: S61978
R;Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S61959
A;Accession: S61978
A;Molecule type: DNA
A;Residues: 1-342 <WAN>
A;Cross-references: EMBL:U43281; NID:g1151218; PIDN:AAB68211.1; PID:g1151238; MIPS:YP
C;Genetics:
A;Map position: 16L
C;Superfamily: conserved hypothetical protein YPL088w

Query Match   65.3%;   Score 32;   DB 2;   Length 342;
Best Local Similarity 70.0%;   Pred. No. 45;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   3;   Gaps   0;

Qy   1 YRLLIRLDER 10
          | ||| |||
Db   207 YNLLYRDER 216

RESULT 11
I37271
cylicin II - human
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 25-Oct-1996
A;Accession: I37271; S52774
R;Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.
Exp. Cell Res. 218, 174-182, 1995
A;Title: The protein complexity of the cytoskeleton of bovine and human sperm heads:
A;Reference number: I37271; MUID:95255491
A;Accession: I37271
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-348 <HES>
A;Cross-references: EMBL:Z46788; NID:g758586; PID:g758587

Query Match   65.3%;   Score 32;   DB 2;   Length 348;
Best Local Similarity 60.0%;   Pred. No. 45;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLDER 10
          | ||:|:||
Db   78 YRSLMRISER 87

RESULT 12
F70548
probable menD protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A;Accession: F70548
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: F70548
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-554 <COL>
A;Cross-references: GB:Z95558; GB:AL123456; NID:g3261781; PID:e316800; PID:g2114017
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: menD
C;Superfamily: menD protein

Query Match   65.3%;   Score 32;   DB 1;   Length 554;
Best Local Similarity   66.7%;   Pred. No. 73;
Matches   6;   Conservative   2;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   2 RLLIRLDER
        || :|:|||
Db   48 RLHVRIDER 56

RESULT 13
B70963
hypothetical protein Rv0236c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C;Accession: B70963
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajndream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70963
A;Molecule type: DNA
A;Residues: 1-1400 <COL>
A;Cross-references: GB:Z92669; GB:AL123456; NID:g3242271; PID:e1300719; PID:g2242273
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv0236c

Query Match   65.3%;   Score 32;   DB 2;   Length 1400;
Best Local Similarity   75.0%;   Pred. No. 1.9e+02;
Matches   6;   Conservative   2;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3 LLIRLDER
        :|:|||||
Db   621 VLLRLDER 628

RESULT 14
S25566
phosphopyruvate hydratase (EC 4.2.1.11) - Streptomyces hygroscopicus (fragment)
C;Species: Streptomyces hygroscopicus
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jul-1997
C;Accession: S25666
R;Hidaka, T.; Hidaka, M.; Uozumi, T.; Seto, H.
Mol. Gen. Genet. 233, 476-478, 1992
A;Title: Nucleotide sequence of a carboxyphosphonoenolpyruvate phosphonomutase gene isol
ns.
A;Reference number: S23585; MUID:92318902
A;Accession: S25666
A;Molecule type: DNA
A;Residues: 1-30 <HID>
A;Cross-references: EMBL:D00609
A;Experimental source: strain SF1293
C;Superfamily: enolase
C;Keywords: carbon-oxygen lyase; gluconeogenesis; glycolysis; hydro-lyase

---

Query Match   63.3%;   Score 31;   DB 2;   Length 30;
Best Local Similarity   66.7%;   Pred. No. 5.8;
Matches   6;   Conservative   1;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLDE 9
       | |:|||
Db   17 YNQLLRLDE 25

RESULT 15
H75054
hypothetical protein PAB2372 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: H75054
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: H75054
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-157 <KAW>
A;Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50333.1; PID:e151
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB2372

Query Match   63.3%;   Score 31;   DB 2;   Length 157;
Best Local Similarity   75.0%;   Pred. No. 32;
Matches   6;   Conservative   1;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 YRLLIRLD 8
       ||||:|:|
Db   137 YRLLLRRD 144

Search completed: February 7, 2000, 11:54:26
Job time: 24336 sec

```
              GenCore version 4.5
       Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:  February 8, 2000, 00:59:53 ; Search time 63.71 Seconds
                                      (without alignments)
                                      4.688 Million cell updates/sec

Title:           US-08-653-294-15
Perfect score:   49
Sequence:        1 YRLLIRLDER 10

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters:      82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :       SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                           SUMMARIES
```

Result No.	Score	% Query Match	DB	Length	ID	Description
1	35	71.4	1	513	RFP_HUMAN	P14373 homo sapien
2	35	71.4	1	522	RFP_MOUSE	Q62158 mus musculu
3	32	65.3	1	299	RL5_BOMMO	O76190 bombyx mori
4	32	65.3	1	317	OLF3_CANFA	Q95156 canis famil
5	32	65.3	1	317	OLF3_HUMAN	Q13607 homo sapien
6	32	65.3	1	348	CYL2_HUMAN	Q14093 homo sapien
7	31	63.3	1	221	YFH8_YEAST	P43592 saccharomyc
8	31	63.3	1	327	RL5_ANOGA	O44248 anopheles g
9	31	63.3	1	627	GIDA_COXBU	P94613 coxiella bu
10	31	63.3	1	825	RCA1_YEAST	P40341 saccharomyc
11	31	63.3	1	1489	YGP0_YEAST	P53115 saccharomyc
12	31	63.3	1	1709	CHD1_HUMAN	O14646 homo sapien
13	31	63.3	1	1711	CHD1_MOUSE	P40201 mus musculu
14	30	61.2	1	85	SPOE_BACSU	P05043 bacillus su
15	30	61.2	1	96	Y112_SSV1	P20219 sulfolobus
16	30	61.2	1	294	RL5A_SCHPO	P52822 schizosacch
17	30	61.2	1	294	RL5B_SCHPO	O74306 schizosacch
18	30	61.2	1	297	RL5_HELAN	O65353 helianthus
19	30	61.2	1	354	VGLI_VZVD	P09258 varicella-z
20	30	61.2	1	404	Y349_MYCGE	P47591 mycoplasma
21	30	61.2	1	595	TRPE_ARATH	P32068 arabidopsis
22	30	61.2	1	611	EMPA_VIBAN	P43147 vibrio angu
23	30	61.2	1	621	GIDA_HELPY	P56138 helicobacte
24	30	61.2	1	621	TRPX_ARATH	P32069 arabidopsis
25	30	61.2	1	628	GIDA_BACSU	P25812 bacillus su
26	30	61.2	1	629	GIDA_HAEIN	P44763 haemophilus
27	30	61.2	1	635	TRG4_ECOLI	Q00185 escherichia
28	30	61.2	1	637	TRG5_ECOLI	Q00184 escherichia
29	30	61.2	1	659	YYBT_BACSU	P37484 bacillus su
30	30	61.2	1	851	STA2_HUMAN	P52630 homo sapien
31	30	61.2	1	864	STA2_PIG	O02799 sus scrofa
32	30	61.2	1	899	SECA_BORBU	O07497 borrelia bu
33	30	61.2	1	3079	IRA2_YEAST	P19158 saccharomyc
34	30	61.2	1	3224	N358_HUMAN	P49792 homo sapien
35	29	59.2	1	101	YEB4_YEAST	P39999 saccharomyc
36	29	59.2	1	124	RL5_PIG	Q95276 sus scrofa
37	29	59.2	1	172	UBC2_YEAST	P06104 saccharomyc
38	29	59.2	1	179	UBC2_CANAL	O74201 candida alb
39	29	59.2	1	293	DBFB_PSEPA	P47243 pseudomonas
40	29	59.2	1	293	RL5_CAEEL	P49405 caenorhabdi
41	29	59.2	1	295	RL5A_XENLA	P15125 xenopus lae
42	29	59.2	1	295	RL5B_XENLA	P15126 xenopus lae
43	29	59.2	1	296	RL5_CHICK	P22451 gallus gall
44	29	59.2	1	296	RL5_HUMAN	P46777 homo sapien
45	29	59.2	1	296	RL5_RAT	P09895 rattus norv

```
                           ALIGNMENTS

RESULT 1
RFP_HUMAN
ID   RFP_HUMAN      STANDARD;      PRT;   513 AA.
AC   P14373;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ZINC-FINGER PROTEIN RFP (RET FINGER PROTEIN).
GN   RFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88246464.
RA   TAKAHASHI M., INAGUMA Y., HIAI H., HIROSE F.;
RT   "Developmentally regulated expression of a human 'finger'-containing
RT   gene encoded by the 5' half of the ret transforming gene.";
RL   Mol. Cell. Biol. 8:1853-1856(1988).
CC   -!- FUNCTION: MAY FUNCTION IN MALE GERM CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- DISEASE: RECOMBINATION OF RFP WITH A PROTEIN
CC       TYROSINE KINASE PRODUCES THE RET TRANSFORMING PROTEIN.
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; J03407; AAA36564.1; -.
DR   PIR; A28101; TVHURF.
DR   MIM; 602165; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
DR   PFAM; PF00622; SPRY; 1.
DR   PFAM; PF00643; zf-B_box; 1.
KW   Proto-oncogene; zinc-finger; Metal-binding; Chromosomal translocation;
KW   Nuclear protein; DNA-binding.
FT   SITE      315    316       BREAKPOINT FOR TRANSLOCATION TO FORM THE
FT                              RFP-RET ONCOGENE.
FT   ZN_FING    16     56       C3HC4-TYPE.
FT   DOMAIN     96    127       B BOX.
SQ   SEQUENCE   513 AA;  58489 MW;  022BC859 CRC32;

Query Match             71.4%;   Score 35;  DB 1;  Length 513;
Best Local Similarity   77.8%;   Pred. No. 8.3;
Matches    7;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     1 YRLLIRLDE 9
           |||| ||:|
Db   198 YRLLARLEE 206
```

```
RESULT  2
RFP_MOUSE
ID   RFP_MOUSE      STANDARD;      PRT;      522 AA.
AC   Q62158;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ZINC-FINGER PROTEIN RFP (RET FINGER PROTEIN).
GN   RFP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 9176437.
RA   CAO T., SHANNON M., HANDEL M.A., ETKIN L.D.;
RT   "Mouse ret finger protein (rfp) proto-oncogene is expressed at
RT   specific stages of mouse spermatogenesis.";
RL   Dev. Genet. 19:309-320(1996).
CC   -!- FUNCTION: MAY FUNCTION IN MALE GERM CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L46855; AAA85354.1; -.
DR   MGD; MGI:97904; RFP.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
DR   PFAM; PF00622; SPRY; 1.
DR   PFAM; PF00643; zf-B_box; 1.
KW   Zinc-finger; Metal-binding; Nuclear protein; DNA-binding.
FT   ZN_FING      25     65       C3HC4-TYPE.
FT   DOMAIN      105    136       B BOX.
SQ   SEQUENCE   522 AA;   59550 MW;   18E6E716 CRC32;

Query Match                71.4%;
Best Local Similarity      77.8%;    Pred. No. 8.5;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    0;

Qy     1 YRLLIRLDE 9
           |||.||:||
Db   207 YRLLARLEE 215

RESULT  3
RL5_BOMMO
ID   RL5_BOMMO      STANDARD;      PRT;      299 AA.
AC   O76190;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L5.
GN   RPL5.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SILK GLAND;
RA   YANG C.S., SEHNAL F.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF008229; AAC24960.1; -.
DR   PFAM; PF00861; Ribosomal_L18p; 1.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   299 AA;   34378 MW;   7262D2FC CRC32;

Query Match                65.3%;
Best Local Similarity      50.0%;    Score 32;   DB 1;   Length 299;
Matches    5;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     1 YRLLIRLDER 10
           |||:||:|| :
Db    49 YRLIVRLSNK 58

RESULT  4
OLF3_CANFA
ID   OLF3_CANFA      STANDARD;      PRT;      317 AA.
AC   Q95156;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   OLFACTORY RECEPTOR-LIKE PROTEIN OLF3.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97008103.
RA   ISSEL-TARVER L., RINE J.;
RT   "Organization and expression of canine olfactory receptor genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:10897-10902(1996).
CC   -!- FUNCTION: PUTATIVE ODORANT OR SPERM CELL RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE RECEPTOR.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U53681; AAB37241.1; -.
DR   GCRDB; GCR_1190; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Olfaction.
FT   DOMAIN        1     25       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     26     49       1 (POTENTIAL).
FT   DOMAIN       50     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     58     79       2 (POTENTIAL).
FT   DOMAIN       80    100       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    101    120       3 (POTENTIAL).
FT   DOMAIN      121    139       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    140    158       4 (POTENTIAL).
FT   DOMAIN      159    196       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    197    219       5 (POTENTIAL).
FT   DOMAIN      220    236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    237    260       6 (POTENTIAL).
FT   DOMAIN      261    272       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM     273    292       7 (POTENTIAL).
FT   DOMAIN       293    317       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD       5            POTENTIAL.
SQ   SEQUENCE     317 AA;   35238 MW;  C16156EE CRC32;

Query Match   Best Local Similarity  65.3%;   Score 32;  DB 1;  Length 317;
                             Conservative 87.5%;  Pred. No. 21;
Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LLIRLDER 10
           ||||| |
Db      47 LLIRLDSR 54

RESULT 5
OLF3_HUMAN
ID   OLF3_HUMAN     STANDARD;      PRT;    317 AA.
AC   Q13607;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   OLFACTORY RECEPTOR-LIKE PROTEIN OLF3.
GN   OLF3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ISSEL-TARVER L., RINE J.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U56421; AAB01215.1; -.
DR   GCRDB; GCR_1924; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Olfaction.
FT   TRANSMEM       1     24       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      25     48       1 (POTENTIAL).
FT   DOMAIN        49     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      58     79       2 (POTENTIAL).
FT   DOMAIN        80    100       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     101    120       3 (POTENTIAL).
FT   DOMAIN       121    139       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     140    160       4 (POTENTIAL).
FT   DOMAIN       161    200       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     201    222       5 (POTENTIAL).
FT   DOMAIN       223    236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     237    261       6 (POTENTIAL).
FT   DOMAIN       262    272       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     273    292       7 (POTENTIAL).
FT   DOMAIN       293    317       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD       5            POTENTIAL.
FT   DISULFID      97    189       BY SIMILARITY.
SQ   SEQUENCE     317 AA;   35316 MW;  30A35A35 CRC32;

Query Match   Best Local Similarity  65.3%;   Score 32;  DB 1;  Length 317;
                             Conservative 87.5%;  Pred. No. 21;
Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       3 LLIRLDER 10
           ||||| |
Db      47 LLIRLDSR 54

RESULT 6
CYL2_HUMAN
ID   CYL2_HUMAN     STANDARD;      PRT;    348 AA.
AC   Q14093;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYLICIN II (MULTIPLE-BAND POLYPEPTIDE II).
GN   CYL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 95255491.
RA   HESS H., HEID H., ZIMBELMANN R., FRANKE W.W.;
RT   "The protein complexity of the cytoskeleton of bovine and human sperm
RT   heads: the identification and characterization of cylicin II.";
RL   Exp. Cell Res. 218:174-182(1995).
CC   -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC       BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z46788; CAA86752.1; -.
KW   Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT   DOMAIN        25    347       31 X 3 AA REPEATS OF K-K-X.
FT   DOMAIN       157    240       3 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT       157    184       1.
FT   REPEAT       185    212       2.
FT   REPEAT       213    240       3.
SQ   SEQUENCE     348 AA;   39079 MW;  FD27FBEF CRC32;

Query Match   Best Local Similarity  65.3%;   Score 32;  DB 1;  Length 348;
                             Conservative 60.0%;  Pred. No. 23;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YRLLIRLDER 10
           || |:|:||
Db      78 YRSLMRISER 87

RESULT 7
YFH8_YEAST
ID   YFH8_YEAST     STANDARD;      PRT;    221 AA.
AC   P43592;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 25.9 KD PROTEIN IN MPR1-GCN20 INTERGENIC REGION.
GN   YFR008W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 95400292.
```

```
RA   MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
RA   SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA   YAMAZAKI M., TASHIRO H., EKI T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; D50617; BAA09247.1; -.
KW   Hypothetical protein.
FT   DOMAIN       11     22       POLY-GLN.
SQ   SEQUENCE   221 AA;  25915 MW;  3D34764A CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 221;
Best Local Similarity   85.7%;  Pred. No. 23;
Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1  YRLLIRL  7
           |||:||
Db    211  YRLLLRL  217


RESULT 8
RL5_ANOGA
ID   RL5_ANOGA      STANDARD;      PRT;    327 AA.
AC   O44248;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L5.
GN   RPL5.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC   Culicidae; Anopheles.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G3;
RA   CORNEL A.J., KUMAR V., MUKABAYIRE O., SALAZAR RAFFERTY C.,
RA   PETRARCA V., COLUZZI M., COLLINS F.H.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS 5S RNA (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AF002238; AAB97731.1; -.
DR   PFAM; PF00861; Ribosomal_L18p; 1.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   327 AA;  37996 MW;  F3A3EED2 CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 327;
Best Local Similarity   50.0%;  Pred. No. 35;
Matches    5;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1  YRLLIRLDER  10
           :|:::||    |
Db     49  FRLIVRLSNR  58
```

```
RESULT 9
GIDA_COXBU
ID   GIDA_COXBU     STANDARD;      PRT;    627 AA.
AC   P94613;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   GLUCOSE INHIBITED DIVISION PROTEIN A.
GN   GIDA.
OS   Coxiella burnetii.
OC   Bacteria; Proteobacteria; gamma subdivision; Coxiella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NINE MILE PHASE I;
RA   WILLEMS H., JAEGER C.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; Y10436; CAA71459.1; -.
DR   PROSITE; PS01280; GIDA_1; 1.
DR   PROSITE; PS01281; GIDA_2; 1.
DR   PFAM; PF01134; GIDA; 1.
SQ   SEQUENCE   627 AA;  69951 MW;  B9AF4071 CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 627;
Best Local Similarity   75.0%;  Pred. No. 70;
Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1  YRLLIRLD  8
           ||||:|  |
Db    436  YRLLRQD  443


RESULT 10
RCA1_YEAST
ID   RCA1_YEAST     STANDARD;      PRT;    825 AA.
AC   P40341;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   MITOCHONDRIAL RESPIRATORY CHAIN COMPLEXES ASSEMBLY PROTEIN RCA1
DE   (EC 3.4.24.-) (TAT-BINDING HOMOLOG 12).
GN   RCA1 OR YTA12 OR YMR089C OR YM9582.14C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014441;
RA   TZAGOLOFF A., YUE J., JANG J., PAUL M.F.;
RT   "A new member of a family of ATPases is essential for assembly of
RT   mitochondrial respiratory chain and ATP synthetase complexes in
RT   Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 269:26144-26151(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 95274317;
RA   SCHNALL R., MANNHAUPT G., STUCKA R., TAUER R., EHNLE S.,
RA   SCHWARZLOSE C., VETTER I., FELDMANN H.;
RT   "Identification of a set of yeast genes coding for a novel family of
```

```
RT   putative ATPases with high similarity to constituents of the 26S
RT   protease complex.";
RL   Yeast 10:1141-1155(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   GENTLES S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
     -!- FUNCTION: PUTATIVE ATP-DEPENDENT PROTEASE. NECESSARY FOR THE
CC       ASSEMBLY OF MITOCHONDRIAL RESPIRATORY CHAIN AND ATPASE COMPLEXES.
CC       FUNCTION BOTH IN POSTTRANSLATIONAL ASSEMBLY AND IN THE TURNOVER OF
CC       MISTRANSLATED OR MISFOLDED POLYPEPTIDES.
CC   -!- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC
CC       METALLOPROTEASE).
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U09358; AAA62606.1; -.
DR   EMBL; X81068; CAA56955.1; -.
DR   EMBL; Z49259; CAA89236.1; -.
DR   SGD; L0002564; YTA12.
DR   PROSITE; PS00674; AAA; 1.
DR   PFAM; PF00004; AAA; 1.
DR   PFAM; PF01434; Peptidase_M41; 1.
KW   ATP-binding; Mitochondrion; Transmembrane; Hydrolase; Metalloprotease;
KW   Zinc.
FT   TRANSMEM    178    194       .POTENTIAL.
FT   TRANSMEM    294    311       .POTENTIAL.
FT   NP_BIND     388    395       ATP (POTENTIAL).
FT   METAL       613    613       ZINC (CATALYTIC).
FT   ACT_SITE    614    614       BY SIMILARITY.
FT   METAL       617    617       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CONFLICT    349    350       DV -> EL (IN REF. 2).
FT   CONFLICT    653    653       I -> V (IN REF. 1).
SQ   SEQUENCE    825 AA;  93276 MW;  69EBD054 CRC32;

Query Match             63.3%;  Score 31; DB 1; Length 825;
Best Local Similarity   77.8%;  Pred. No. 93;
Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YRLLIRLDE 9
            |||| || |
Db       36 YRLLNRLQE 44

RESULT 11
YGP0_YEAST
ID   YGP0_YEAST     STANDARD;      PRT;   1489 AA.
AC   P53115;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 171.5 KD HELICASE IN NUT1-ARO2 INTERGENIC REGION.
GN   YGL150C OR G1880.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96158061.
RA   JAMES C.M., INDGE K.J., OLIVER S.G.;
RT   "DNA sequence analysis of a 35 kb segment from Saccharomyces
RT   cerevisiae chromosome VII reveals 19 open reading frames including
```

```
RT   RAD54, ACE1/CUP2, PMR1, RCK1, AMS1 and CAL1/CDC43.";
RL   Yeast 11:1413-1419(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1769;
RX   MEDLINE; 97197983;
RA   VOET M., DEFOOR E., VERHASSELT P., RILES L., ROBBEN J., VOLCKAERT G.;
RT   "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT   chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT   TIP1, MRF1 genes and six new open reading frames.";
RL   Yeast 13:177-182(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z48618; CAA88537.1; -.
DR   EMBL; Z72672; CAA96861.1; -.
DR   EMBL; X99960; CAA68224.1; -.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
KW   Hypothetical protein; Nuclear protein; DNA-binding; Helicase;
KW   ATP-binding.
FT   DOMAIN      188    193       POLY-ALA.
FT   DOMAIN      259    268       POLY-GLU.
FT   DOMAIN      300    306       POLY-SER.
FT   DOMAIN      568    573       POLY-GLU.
FT   DOMAIN      675    682       POLY-GLU.
FT   NP_BIND     731    738       ATP (POTENTIAL).
FT   SITE        841    844       DEAQ BOX.
SQ   SEQUENCE   1489 AA;  171454 MW;  8149887E CRC32;

Query Match             63.3%;  Score 31; DB 1; Length 1489;
Best Local Similarity   58.3%;  Pred. No. 1.7e+02;
Matches    7;  Conservative    3;  Mismatches    0;  Indels    2;  Gaps    1;

Qy        1 YRLLIR--LDER 10
            ||||| :  :||
Db     1420 YRLLVRGTIEER 1431

RESULT 12
CHD1_HUMAN
ID   CHD1_HUMAN     STANDARD;      PRT;   1709 AA.
AC   O14646;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1 (CHD-1).
GN   CHD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97470991.
RA   WOODAGE T., BASRAI M.A., BAXEVANIS A.D., HIETER P., COLLINS F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC       IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------
DR   EMBL; AF006513; AAB87381.1; -.
DR   HSSP; P23197; 1AP0.
DR   MIM; 602118; -.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00385; chromo; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN       1     70    SER-RICH.
FT   DOMAIN     117    137    SER-RICH.
FT   DOMAIN     272    364    CHROMO DOMAIN.
FT   DOMAIN     389    452    CHROMO DOMAIN.
FT   NP_BIND    506    513    ATP (POTENTIAL).
FT   SITE       614    617    DEAH BOX.
FT   DOMAIN    1628   1644    3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT    1628   1632    1.
FT   REPEAT    1634   1638    2.
FT   REPEAT    1640   1644    3.
SQ   SEQUENCE  1709 AA;  196517 MW;  EC7F932A CRC32;

Query Match             63.3%;      Score 31;  DB 1;  Length 1709;
Best Local Similarity   77.8%;      Pred. No. 2e+02;
Matches    7;  Conservative    7;  Mismatches    1;  Indels    1;  Gaps    0;

Qy      2 RLLIRLDER 10
          :|||| ||
Db    795 KLLIRLRER 803

RESULT 13
CHD1_MOUSE
ID   CHD1_MOUSE     STANDARD;      PRT;   1711 AA.
AC   P40201;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1 (CHD-1).
GN   CHD1 OR CHD-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93211972.
RA   DELMAS V., STOKES D.G., PERRY R.P.;
RT   "A mammalian DNA-binding protein that contains a chromodomain and an
RT   SNF2/SWI2-like helicase domain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
CC       IMPORTANT ROLE IN GENE REGULATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: ABUNDANCE IS HIGHER IN CELLS REPRESENTING
CC       EARLY STAGES OF THE B LYMPHOID LINEAGE SUCH AS PRE-B AND B CELLS,
CC       THAN IN CELLS REPRESENTING MATURE PLASMACYTES OR OTHER CELL
CC       LINEAGES SUCH AS FIBROBLASTS.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   -------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------
DR   EMBL; L10410; AAB08486.1; -.
DR   PIR; A47392; A47392.
DR   HSSP; P23197; 1AP0.
DR   MGD; MGI:88393; CHD1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
DR   PFAM; PF00176; SNF2_N; 1.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00385; chromo; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN       1     70    SER-RICH.
FT   DOMAIN     116    136    SER-RICH.
FT   DOMAIN     270    362    CHROMO DOMAIN.
FT   DOMAIN     387    450    CHROMO DOMAIN.
FT   NP_BIND    504    511    ATP (POTENTIAL).
FT   SITE       612    615    DEAH BOX.
FT   DOMAIN    1629   1645    3 X 5 AA REPEATS OF H-S-D-H-R.
FT   REPEAT    1629   1633    1.
FT   REPEAT    1635   1639    2.
FT   REPEAT    1641   1645    3.
SQ   SEQUENCE  1711 AA;  196409 MW;  CB184D33 CRC32;

Query Match             63.3%;      Score 31;  DB 1;  Length 1711;
Best Local Similarity   77.8%;      Pred. No. 2e+02;
Matches    7;  Conservative    7;  Mismatches    1;  Indels    1;  Gaps    0;

Qy      2 RLLIRLDER 10
          :|||| ||
Db    793 KLLIRLRER 801

RESULT 14
SPOE_BACSU
ID   SPOE_BACSU     STANDARD;      PRT;     85 AA.
AC   P05043;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   STAGE 0 SPORULATION REGULATORY PROTEIN.
GN   SPO0E.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 88260878.
RA   PEREGO M., HOCH J.A.;
RT   "Isolation and sequence of the spo0E gene: its role in initiation of
RT   sporulation in Bacillus subtilis.";
RL   Mol. Microbiol. 1:125-132(1987).
CC   -!- MISCELLANEOUS: MUTATIONS THAT BLOCK THE ONSET OF SPORULATION IN
CC       BACILLUS SUBTILIS ARE CALLED STAGE 0 MUTATIONS, OR SPO0 MUTATIONS,
CC       AND AT LEAST EIGHT GENETIC LOCI HAVE BEEN IDENTIFIED IN WHICH A
CC       MUTATION CAN RESULT IN A STAGE 0 PHENOTYPE. STAGE 0 MUTANTS LACK
CC       THE ABILITY TO FORM THE ASYMETRIC SEPTUM CHARACTERISTIC OF THE
CC       INITIATION OF THE SPORULATION PROCESS. IN ADDITION, THESE
CC       MUTATIONS HAVE MANY PLEIOTROPIC EFFECTS ON THE SYNTHESIS OF A
CC       VARIETY OF TRANSCRIPTS BY MINOR FORMS OF RNA POLYMERASE IN THIS
CC       ORGANISM.
CC   -------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Y00526; CAA68583.1; -.
```

```
DR   EMBL; Z99111; CAB13237.1; -.
DR   PIR; S03746; S03746.
DR   SUBTILIST; BG10769; SPO0E.
KW   Sporulation;Transcription regulation.
SQ   SEQUENCE   85 AA;  9791 MW;  E2B23676 CRC32;

Query Match             61.2%;  Score 30;  DB 1;  Length 85;
Best Local Similarity   55.6%;  Pred. No. 14;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 RLLIRLDER 10
          ||:::||:
Db      9 RLLVSIDEK 17


RESULT 15
Y112_SSV1
ID   Y112_SSV1      STANDARD;      PRT;       96 AA.
AC   P20219;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   HYPOTHETICAL 11.2 KD PROTEIN (ORF E-96).
OS   Sulfolobus virus-like particle SSV1.
OC   Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92024080.
RA   PALM P., SCHLEPER C., GRAMPP B., YEATS S., MCWILLIAM P., REITER W.-D.,
RA   ZILLIG W.;
RT   "Complete nucleotide sequence of the virus SSV1 of the
RT   archaebacterium Sulfolobus shibatae.";
RL   Virology 185:242-250(1991).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X07234; CAA30218.1; -.
DR   PIR; S03219; S03219.
KW   Hypothetical protein.
SQ   SEQUENCE   96 AA;  11176 MW;  D351EB9B CRC32;

Query Match             61.2%;  Score 30;  DB 1;  Length 96;
Best Local Similarity   44.4%;  Pred. No. 16;
  Matches    4;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 YRLLIRLDE 9
          |:::::| ||
Db     75 YKIILRCDE 83


Search completed: February  8, 2000, 00:59:54
Job time: 3783 sec
```

OM of: US-08-653-294-13 to: EST:*   out_format : pfs
Date: Feb 8, 2000   4:02 AM

About: Results were produced by the GenCore software, version 4.5,
      Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160700_15770/app_query.fasta.1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-13
Query length: 10
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 8553.360000

score_list:

sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_gss11:AQ324186	+	39.00	142.06	37.98	784	! AQ324186 mgxb0017A14r CUGI Rice
gb_est5:D75467	+	37.00	141.45	41.05	335	! D75467 CELK104H2F Yuji Kohara u
gb_est5:D73508	+	37.00	140.81	44.57	360	! D73508 CELK051H2F Yuji Kohara u
gb_est5:D73590	+	37.00	140.81	44.57	360	! D73590 CELK066G3F Yuji Kohara u
gb_est5:D74728	+	37.00	140.81	44.57	360	! D74728 CELK084F3F Yuji Kohara u
gb_est5:D75423	+	37.00	140.81	44.57	360	! D75423 CELK104C7F Yuji Kohara u
gb_est5:D75727	+	37.00	140.81	44.57	360	! D75727 CELK109D2F Yuji Kohara u
gb_est16:C40575	+	37.00	140.81	44.57	360	! C40575 C40575 Yuji Kohara unpub
gb_est16:C42212	+	37.00	140.81	44.57	360	! C42212 C42212 Yuji Kohara unpub
gb_est16:C44981	+	37.00	140.81	44.57	360	! C44981 C44981 Yuji Kohara unpub
gb_est16:C51139	+	37.00	140.81	44.57	360	! C51139 C51139 Yuji Kohara unpub
gb_est36:AV192603	+	37.00	140.81	44.57	360	! AV192603 AV192603 Yuji Kohara u
gb_est36:AV192774	+	37.00	140.81	44.57	360	! AV192774 AV192774 Yuji Kohara u
gb_est36:AV193922	+	37.00	140.81	44.57	360	! AV193922 AV193922 Yuji Kohara u
gb_est36:AV194609	+	37.00	140.81	44.57	360	! AV194609 AV194609 Yuji Kohara u
gb_est36:AV191482	+	37.00	140.57	45.99	370	! AV191482 AV191482 Yuji Kohara u
gb_est16:C39480	+	37.00	140.47	46.56	374	! C39480 C39480 Yuji Kohara unpub
gb_est16:C42649	+	37.00	140.47	46.56	374	! C42649 C42649 Yuji Kohara unpub
gb_est16:C48686	+	37.00	140.42	46.84	376	! C48686 C48686 Yuji Kohara unpub
gb_est16:C43167	+	37.00	140.40	46.98	377	! C43167 C43167 Yuji Kohara unpub
gb_est16:AV191244	+	37.00	140.40	47.13	378	! AV191244 AV191244 Yuji Kohara u
gb_est28:AI531870	+	37.00	138.72	58.25	455	! AI531870 SD03153.5prime SD Dros
gb_gss8:AQ038578	+	37.00	138.38	60.90	473	! AQ038578 CIT-HSP-2333K12.TR CIT
gb_gss12:AA280470	-	36.00	136.87	73.90	354	! AA280470 SWMFCA2019SK Brugia ma
gb_gss9:AQ146935	-	36.00	134.63	98.47	455	! AQ146935 HS_2248_A2_B10_MR CIT
gb_gss8:AQ038143	-	35.00	132.77	124.93	354	! AQ038143 CIT-HSP-2320M23.TR CII
gb_gss12:AQ331915	-	35.00	132.33	132.22	372	! AQ331915 HS_5002_A2_D11_SP6E RP
gb_est36:AV189109	+	35.00	132.26	133.44	375	! AV189109 AV189109 Yuji Kohara u
gb_est17:C73120	-	35.00	130.44	168.55	460	! C73120 C73120 Rice panicle at f
gb_est23:AU021869	-	35.00	129.73	184.56	498	! AU021869 AU021869 Mouse unferti
gb_gss11:AQ294280	+	35.00	129.71	184.98	499	! AQ294280 HS_3000_B1_F08_MR CIT
gb_gss14:AQ543593	+	35.00	129.68	185.83	501	! AQ543593 RPCI-11-359C3.TV RPCI-
gb_est44:AV390501	+	35.00	129.32	205.46	547	! AV390501 AV390501 Rice callus C
gb_est32:AU068481	-	34.00	132.00	137.99	244	! AU068481 AU068481 Rice callus O
gb_est44:AU081183	-	34.00	130.01	178.10	305	! AU081183 AU081183 Oncorhynchus
gb_est1:D36382	-	34.00	129.31	194.89	330	! D36382 CELK032E2F Yuji Kohara u
gb_gss11:AQ281049	+	34.00	128.78	208.45	350	! AQ281049 RPCI11-7639.TJ RPCI-11
gb_est2:T98600	-	34.00	128.68	211.18	354	! T98600 ye60e09.s1 Soares fetal
gb_est1:D36749	-	34.00	128.53	215.27	360	! D36749 CELK036F4F Yuji Kohara u
gb_est36:AV201748	-	34.00	128.53	215.27	360	! AV201748 AV201748 Yuji Kohara u
gb_est17:C63035	-	34.00	128.19	224.87	374	! C63035 C63035 Yuji Kohara unpub
gb_est8:C08467	+	34.00	128.07	228.31	379	! C08467 C08467 Yuji Kohara unpub

gb_est9:C13188	-	34.00	128.05	229.00	380	! C13188 C13188 Yuji Kohara un
gb_gss:AQ128909	-	34.00	127.82	235.95	390	! AQ128909 HS_3028_B2_C12_MR C
gb_gss6:AQ843768	+	34.00	127.24	253.97	416	! AQ843768 LMAJFV1_1m02c05.x3

seq_name: gb_gss11:AQ324186

seq_documentation_block:
LOCUS   AQ324186   784 bp   DNA   GSS   08-JAN-1999
DEFINITION   mgxb0017A14r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
           clone mgxb0017A14r, genomic survey sequence.
ACCESSION   AQ324186
VERSION   AQ324186.1   GI:4116038
KEYWORDS   GSS.
SOURCE   Magnaporthe grisea
  ORGANISM   Magnaporthe grisea
          Unclassified.
REFERENCE   1 (bases 1 to 784)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
          Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE   A BAC End Sequencing Framework to sequence the Magnaporthe grisea
          Genome
  JOURNAL   Unpublished (1998)
COMMENT   Contact: Dean RA
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson University, Clemson, SC 29634
          Tel: 864 656 5737
          Fax: 864 656 4293
          Email: rdean@clemson.edu
          Seq primer: GGAAACAGCTATGACCATG
          Class: BAC ends
          High quality sequence start: 95
          High quality sequence stop: 334.
FEATURES   Location/Qualifiers
    source   1..784
          /organism="Magnaporthe grisea"
          /strain="70-15"
          /clone="mgxb0017A14r"
          /clone_lib="CUGI Rice Blast BAC Library"
          /tissue_type="Protoplasts"
          /lab_host="E. coli DH10B"
          /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
          Rice blast is one of the most devastating fungal diseases
          of rice world wide. It is a filamentous ascomycete with
          a haploid genome (n=7) of approximately 40 Mbp. Rice
          blast is an important model fungal pathogen for studying
          numerous aspects of the fungal-host interaction. In
          order to facilitate genome wide analysis, a BAC library
          containing 9216 clones with an average insert size of 130
          kbp was constructed. This library represents greater
          than 25X genome coverage. High density colony filters
          are available upon request."
BASE COUNT   208 a   213 c   156 g   206 t   1 others
ORIGIN

alignment_scores:
     Quality: 39.00   Length: 10
     Ratio: 3.900   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-13 x AQ324186/rev

Align seg 1/1 to reverse of: AQ324186   from: 1   to: 784

    1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     |||||||||:||:::.:||:||||||||||
  108 TATAGATTAAGTGTACAGATGGACGAACGG 79

seq_name: gb_est5:D75467

seq_documentation_block:

```
LOCUS       D75467        335 bp    mRNA           EST      18-OCT-1999
DEFINITION  CELK104H2F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk104h2 5', mRNA sequence.
ACCESSION   D75467
VERSION     D75467.1  GI:1121251
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 335)
.AUTHORS    Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp
            Insert Length: 651   Std Error: 0.00.
            Percent Similarity: 90.000   Percent Identity: 70.000
FEATURES    Location/Qualifiers
     source 1..335
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk104h2"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT    94 a    84 c    74 g    80 t    3 others
ORIGIN

alignment_scores:
     Quality:        37.00          Length:        10
     Ratio:          4.111          Gaps:           0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x D75467

Align seg 1/1  to: D75467  from: 1  to: 335

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       |||  |||::||||:::|||||||||||||
   260 TATTCTCTTGCATTCGGATGGATGAACGC 289

seq_name: gb_est5:D73508

seq_documentation_block:
LOCUS       D73508        360 bp    mRNA           EST      18-OCT-1999
DEFINITION  CELK05H2F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk51h2 5', mRNA sequence.
ACCESSION   D73508
VERSION     D73508.1  GI:1119294
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
            On Apr 14, 1993 this sequence version replaced gi:785918.
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp
            Insert Length: 2140   Std Error: 0.00.
            High quality sequence stop: 258.
FEATURES    Location/Qualifiers
     source 1..360
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk51h2"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT   108 a    88 c    83 g    79 t    2 others
ORIGIN

alignment_scores:
     Quality:        37.00          Length:        10
     Ratio:          4.111          Gaps:           0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x D73508

Align seg 1/1  to: D73508  from: 1  to: 360

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       |||  |||::||||:::|||||||||||||
   241 TATTCTCTTGCATTCGGATGGATGAACGC 270

seq_name: gb_est5:D73590

seq_documentation_block:
LOCUS       D73590        360 bp    mRNA           EST      18-OCT-1999
DEFINITION  CELK060G3F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk60g3 5', mRNA sequence.
ACCESSION   D73590
VERSION     D73590.1  GI:1119366
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
            On Apr 14, 1993 this sequence version replaced gi:837454.
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp
            Insert Length: 743   Std Error: 0.00.
            High quality sequence stop: 257.
FEATURES    Location/Qualifiers
     source 1..360
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk60g3"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT   103 a    87 c    84 g    80 t    6 others
```

```
ORIGIN

alignment_scores:
     Quality:      37.00                    Length:        10
     Ratio:         4.111                    Gaps:           0
Percent Similarity:  90.000      Percent Identity:  70.000

alignment_block:
US-08-653-294-13 x D73590        ..

Align seg 1/1   to: D73590   from: 1  to: 360

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||| |||::|||||::|||||||||
   258 TATTCTCTTGCATTCGGATGGATGAACGC 287

seq_name: gb_est5:D74728

seq_documentation_block:
LOCUS       D74728      360 bp    mRNA              EST       18-OCT-1999
DEFINITION  CELK084F3F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk84f3 5', mRNA sequence.
ACCESSION   D74728
VERSION     D74728.1  GI:1120512
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276173.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp
            High quality sequence stop: 296.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk84f3"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT  105 a    88 c    85 g    79 t    3 others
ORIGIN

alignment_scores:
     Quality:      37.00                    Length:        10
     Ratio:         4.111                    Gaps:           0
Percent Similarity:  90.000      Percent Identity:  70.000

alignment_block:
US-08-653-294-13 x D74728        ..

Align seg 1/1   to: D74728   from: 1  to: 360

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||| |||::|||||::|||||||||
   257 TATTCTCTNTGCATTCGGATGGATGAACGC 286

seq_name: gb_est5:D75423

seq_documentation_block:
LOCUS       D75423      360 bp    mRNA              EST       18-OCT-1999
DEFINITION  CELK104C7F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk104c7 5', mRNA sequence.
ACCESSION   D75423
VERSION     D75423.1  GI:1121207
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp
            High quality sequence stop: 120.
            Insert Length: 1013   Std Error: 0.00
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk104c7"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT  106 a    87 c    83 g    79 t    5 others
ORIGIN

alignment_scores:
     Quality:      37.00                    Length:        10
     Ratio:         4.111                    Gaps:           0
Percent Similarity:  90.000      Percent Identity:  70.000

alignment_block:
US-08-653-294-13 x D75423        ..

Align seg 1/1   to: D75423   from: 1  to: 360

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||| |||::|||||::|||||||||
   248 TATTCTCTNTGCATTCGGATGGATGAACGC 277

seq_name: gb_est5:D75727

seq_documentation_block:
LOCUS       D75727      360 bp    mRNA              EST       18-OCT-1999
DEFINITION  CELK109D2F Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk109d2 5', mRNA sequence.
ACCESSION   D75727
VERSION     D75727.1  GI:1121511
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
COMMENT     Contact: Yuji Kohara
```

```
                    Gene Library Lab
                    National Institute of Genetics
                    Yata 1111, Mishima, Shizuoka 411, Japan
                    Tel: 81-559-81-6854
                    Fax: 81-559-81-6855
                    Email: ykohara@lab.nig.ac.jp
                    Insert Length: 1157    Std Error: 0.00.
FEATURES            Location/Qualifiers
     source         1..360
                    /organism="Caenorhabditis elegans"
                    /strain="N2"
                    /db_xref="taxon:6239"
                    /clone="yk109d2"
                    /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                    hermaphrodite embryo"
                    /sex="hermaphrodite"
                    /dev_stage="embryo"
BASE COUNT      105 a     87 c     85 g     75 t       8 others
ORIGIN

alignment_scores:
          Quality:      37.00               Length:       10
          Ratio:       4.111                Gaps:          0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x D75727      ..

Align seg 1/1  to: D75727   from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||:::|||||||||||||
 257 TATTCTCTNTGCATTCGGATGGATGAACGC 286

seq_name: gb_est16:C40575

seq_documentation_block:
LOCUS       C40575          360 bp   mRNA            EST     18-OCT-1999
DEFINITION  C40575 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk246c11 5', mRNA sequence.
ACCESSION   C40575
VERSION     C40575.1  GI:2376812
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata.A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395034.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES            Location/Qualifiers
     source         1..360
                    /organism="Caenorhabditis elegans"
                    /strain="N2"
                    /db_xref="taxon:6239"
                    /clone="yk246c11"
                    /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                    hermaphrodite embryo"
                    /sex="hermaphrodite"
                    /dev_stage="embryo"
BASE COUNT      109 a     93 c     82 g     75 t       1 others
ORIGIN
```

```
          Quality:      37.00               Length:       10
          Ratio:       4.111                Gaps:          0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x C40575      ..

Align seg 1/1  to: C40575   from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||:::|||||||||||||
 214 TATTCTCTTGCATTCGGATGGATGAACGC 243

seq_name: gb_est16:C42212

seq_documentation_block:
LOCUS       C42212          360 bp   mRNA            EST     18-OCT-1999
DEFINITION  C42212 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk290d10 5', mRNA sequence.
ACCESSION   C42212
VERSION     C42212.1  GI:2378449
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata.A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400867.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES            Location/Qualifiers
     source         1..360
                    /organism="Caenorhabditis elegans"
                    /strain="N2"
                    /db_xref="taxon:6239"
                    /clone="yk290d10"
                    /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                    hermaphrodite embryo"
                    /sex="hermaphrodite"
                    /dev_stage="embryo"
BASE COUNT      109 a     90 c     82 g     78 t       1 others
ORIGIN

alignment_scores:
          Quality:      37.00               Length:       10
          Ratio:       4.111                Gaps:          0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x C42212      ..

Align seg 1/1  to: C42212   from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||:::|||||||||||||
 222 TATTCTCTTGCATTCGGATGGATGAACGC 251

seq_name: gb_est16:C44981

seq_documentation_block:
```

```
LOCUS       C44981          360 bp    mRNA            EST       18-OCT-1999
DEFINITION  C44981 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk373f5 5', mRNA sequence.
ACCESSION   C44981
VERSION     C44981.1  GI:2381218
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
TITLE       Expression map of the C.elegans genome
JOURNAL     Unpublished (1996)
COMMENT     On May 8, 1995 this sequence version replaced gi:801522.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk373f5"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT    107 a     95 c     79 g     75 t      4 others
ORIGIN

alignment_scores:
        Quality:      37.00              Length:      10
        Ratio:        4.111              Gaps:        0
Percent Similarity:   90.000   Percent Identity:   70.000

alignment_block:
US-08-653-294-13 x C44981                ..

Align seg 1/1 to: C44981 from: 1  to: 360

        1 TyrArgLeuAlaIleArgLeuAspGluArg 10
          ||| |||::::||||||::::||||||||||
      210 TATTCTCTTGCATTCGGATGGATGAACGC 239

seq_name: gb_est16:C51139

seq_documentation_block:
LOCUS       C51139          360 bp    mRNA            EST       18-OCT-1999
DEFINITION  C51139 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk491f7 5', mRNA sequence.
ACCESSION   C51139
VERSION     C51139.1  GI:2388392
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
TITLE       Expression map of the C.elegans genome
JOURNAL     Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395322.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk491f7"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT    112 a     91 c     82 g     74 t      1 others
ORIGIN

alignment_scores:
        Quality:      37.00              Length:      10
        Ratio:        4.111              Gaps:        0
Percent Similarity:   90.000   Percent Identity:   70.000

alignment_block:
US-08-653-294-13 x C51139                ..

Align seg 1/1 to: C51139 from: 1  to: 360

        1 TyrArgLeuAlaIleArgLeuAspGluArg 10
          ||| |||::::||||||::::||||||||||
      219 TATTCTCTTGCATTCGGATGGATGAACGC 248

seq_name: gb_est36:AV192603

seq_documentation_block:
LOCUS       AV192603        360 bp    mRNA            EST       22-JUL-1999
DEFINITION  AV192603 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk607d3 5', mRNA sequence.
ACCESSION   AV192603
VERSION     AV192603.1  GI:5574755
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
AUTHORS     Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
TITLE       Expressed genes in C.elegans
JOURNAL     Unpublished (1999)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247424.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk607d3"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT    106 a     88 c     81 g     81 t      4 others
ORIGIN
```

```
alignment_scores:
        Quality:    37.00              Length:      10
        Ratio:       4.111             Gaps:         0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x AV192603

Align seg 1/1  to: AV192603  from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||||::||||||||||
 245 TATTCTCTTGCATTCGGATGGATGAACGC 274

seq_name: gb_est36:AV192774

seq_documentation_block:
LOCUS       AV192774       360 bp    mRNA           EST    22-JUL-1999
DEFINITION  AV192774 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk609e10 5', mRNA
            sequence.
ACCESSION   AV192774
VERSION     AV192774.1  GI:5574926
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., and
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y., and
            Nomoto,H.
TITLE       Expressed genes in C.elegans
JOURNAL     Unpublished (1999)
COMMENT     On Feb 18, 1999 this sequence version replaced gi:4299301.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES    Location/Qualifiers
   source   1..360
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk609e10"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT  107 a     88 c    84 g    80 t    1 others
ORIGIN

alignment_scores:
        Quality:    37.00              Length:      10
        Ratio:       4.111             Gaps:         0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x AV192774

Align seg 1/1  to: AV192774  from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||||::||||||||||
 242 TATTCTCTTGCATTCGGATGGATGAACGC 271

seq_name: gb_est36:AV193922


seq_documentation_block:
LOCUS       AV193922       360 bp    mRNA           EST    22-JUL-1999
DEFINITION  AV193922 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk623b7 5', mRNA sequence.
ACCESSION   AV193922
VERSION     AV193922.1  GI:5576074
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y., and
            Nomoto,H.
TITLE       Expressed genes in C.elegans
JOURNAL     Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948755.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES    Location/Qualifiers
   source   1..360
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk623b7"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT  109 a     92 c    82 g    76 t    1 others
ORIGIN

alignment_scores:
        Quality:    37.00              Length:      10
        Ratio:       4.111             Gaps:         0
Percent Similarity:  90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x AV193922

Align seg 1/1  to: AV193922  from: 1  to: 360

   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||| |||::|||||||::||||||||||
 216 TATTCTCTTGCATTCGGATGGATGAACGC 245

seq_name: gb_est36:AV194609

seq_documentation_block:
LOCUS       AV194609       360 bp    mRNA           EST    22-JUL-1999
DEFINITION  AV194609 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk631f9 5', mRNA sequence.
ACCESSION   AV194609
VERSION     AV194609.1  GI:5576761
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 360)
AUTHORS     Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y., and
            Nomoto,H.
```

us-08-653-294-13.rst

TITLE   Expressed genes in C.elegans
JOURNAL   Unpublished (1999)
COMMENT   On May 9, 1996 this sequence version replaced gi:1132813.
           Contact: Yuji Kohara
           Gene Library Lab
           National Institute of Genetics
           Yata 1111, Mishima, Shizuoka 411, Japan
           Tel: 81-559-81-6854
           Fax: 81-559-81-6855
           Email: ykohara@lab.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..360
                    /organism="Caenorhabditis elegans"
                    /strain="N2"
                    /db_xref="taxon:6239"
                    /clone="yk63if9"
                    /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                    hermaphrodite embryo"
                    /sex="hermaphrodite"
                    /dev_stage="embryo"
BASE COUNT   107 a   88 c   83 g   81 t   1 others
ORIGIN

alignment_scores:
       Quality: 37.00   Length: 10
        Ratio: 4.111   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x AV194609   ..

Align seg 1/1 to: AV194609 from: 1 to: 360

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||| |||::::|||||:::||||||||
   239 TATTCTCTTTGCATTCGGATGGATGAACGC 268

```
                 GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 8, 2000, 01:29:38 ; Search time 122.56 Seconds
                                    (without alignments)
                                    1.933 Million cell updates/sec

Title:          US-08-653-294-14
Perfect score:  49
Sequence:       1 REDLRILLRY 10

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:       188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :      A_Geneseq_36:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	10	1	W47271	Immunomodulatory p
2	44	89.8	10	1	R41212	Peptide fragment o
3	44	89.8	10	1	R83075	HLA-B2702 CTL modu
4	44	89.8	10	1	R83094	HLA-B2702 CTL modu
5	44	89.8	10	1	R83096	HLA-B2702 CTL modu
6	44	89.8	10	1	R95423	HLA-B2705.75-84. C
7	44	89.8	10	1	R95425	HLA-B2702.75-84(D)
8	44	89.8	10	1	W07513	T-cell modulating
9	44	89.8	10	1	W47267	Immunomodulatory p
10	44	89.8	10	1	W47269	Immunomodulatory p
11	44	89.8	10	1	W33785	Peptide B2702.75-8
12	44	89.8	10	1	W33787	Peptide B2702.75-8
13	44	89.8	10	1	W33789	Peptide B2702.75-8
14	44	89.8	17	1	R71442	Human HLA-B27-(62-
15	44	89.8	17	1	R71443	Human [Phe74]-HLA-
16	44	89.8	25	1	R41221	Peptide fragment o
17	44	89.8	25	1	R83091	HLA-B2702 CTL modu
18	44	89.8	25	1	R95417	HLA-B2705.60-84. C
19	44	89.8	337	1	P70590	Sequence of the hu
20	44	89.8	362	1	P70155	Sequence encoded b
21	39	79.6	10	1	R41208	Peptide fragment o
22	39	79.6	10	1	R83062	HLA-B2702 CTL modu
23	39	79.6	10	1	R95413	Alpha1-helix of HL
24	39	79.6	10	1	R95427	HLA-B2702.75-84(L)
25	39	79.6	10	1	W07512	T-cell modulating
26	39	79.6	10	1	W07514	T-cell modulating
27	39	79.6	10	1	W47265	Immunomodulatory p
28	39	79.6	10	1	W33784	Peptide B2702.75-8
29	39	79.6	15	1	R92912	HLA-B2702 CTL modu
30	39	79.6	15	1	W33795	Peptide B2702.70-8
31	39	79.6	20	1	R92907	HLA-B2702 CTL modu
32	39	79.6	20	1	R92908	HLA-B2702 CTL modu
33	39	79.6	20	1	R95428	HLA-B2702 84-75-84
34	39	79.6	20	1	W33778	Immunomodulating d
35	39	79.6	20	1	W33791	Peptide B2702.84-7
36	39	79.6	25	1	R41205	Peptide fragment o
37	39	79.6	25	1	R48286	Peptide fragment o
38	39	79.6	25	1	R83090	HLA-B2702 CTL modu
39	39	79.6	25	1	R83093	HLAB38 CTL modulat
40	39	79.6	25	1	R95416	HLA-B2702.60-84. C
41	39	79.6	25	1	R95422	HLAB38.6084. Comps
42	39	79.6	25	1	W33794	Peptide B2702.60-8
43	39	79.6	184	1	Y06801	Peptide Seq ID No:
44	39	79.6	362	1	R03142	Sequence of HLA-Bw
45	39	79.6	362	1	R03144	Sequence of HLA-B5

```
                                ALIGNMENTS

RESULT  1
W47271
ID   W47271 standard; peptide; 10 AA.
AC   W47271;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   Misc_difference 1..10
FT                   /note= "at least one of the amino acids is the
FT                   D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence 10 AA;

        Query Match             100.0%;  Score 49; DB 1; Length 10;
        Best Local Similarity   100.0%;  Pred. No. 0.0009;
        Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 REDLRILLRY 10
              ||||||||||
Db          1 REDLRILLRY 10


RESULT  2
R41212
ID   R41212 standard; peptide; 10 AA.
AC   R41212;
DT   15-MAR-1994 (first entry)
DE   Peptide fragment of Class I HLA peptide.
KW   Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;
KW   parasitic disease; cytotoxic T lymphocyte; modulation.
OS   Synthetic.
PN   WO9317699-A.
PD   16-SEP-1993.
PF   25-FEB-1993; U01758.
```

```
PR   02-MAR-1992; US-844716.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger CA, Krensky AM;
DR   WPI: 93-303134/38.
PT   New peptide(s) based on Class I HLA antigen domains - used for
PT   modulating cytotoxic T-lymphocyte activity towards targets
PS   Claim 11; Page 54; 61pp; English.
CC   The peptide is used to modulate cytotoxic T-lymphocyte (CTL)
CC   activity, either by inhibition or stimulation. It can be used
CC   for inhibiting CTL toxicity in transplantations, for inducing CTL
CC   activity in parasitic diseases and neoplasia and in studies on viral
CC   infection. The peptide can also be used for identifying CTLs which
CC   bind to it and removing subsets of CTLs from a T-cell composition.
CC   This peptide sequence is more commonly found within larger peptide
CC   compounds of not more than 30 amino acids in length.
SQ   Sequence  10 AA;

Query Match           89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity 90.0%;  Pred. No. 0.0084;
Matches   9;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 REDLRILLRY 10
          ||||| ||||
Db    1 REDLRTLLRY 10

RESULT  3
R83075
ID   R83075 standard; peptide; 10 AA.
AC   R83075;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 14; Page 34; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   corresponds to residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence  10 AA;

Query Match           89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity 90.0%;  Pred. No. 0.0084;
Matches   9;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 REDLRILLRY 10
          ||||| ||||
Db    1 REDLRTLLRY.10

RESULT  4
R83094
ID   R83094 standard; peptide; 10 AA.
AC   R83094;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.75-84(D)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 14; Page 34; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   corresponds to residues 75-84 of the alpha-1 domain of the class I MHC
CC   HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with
CC   a subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence  10 AA;

Query Match           89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity 90.0%;  Pred. No. 0.0084;
Matches   9;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 REDLRILLRY 10
          ||||| |||
Db    1 REDLRIALRY 10

RESULT  5
R83096
ID   R83096 standard; peptide; 10 AA.
AC   R83096;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.75-84(L)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 14; Page 34; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   corresponds to residues 75-84 of the alpha-1 domain of the class I MHC
CC   HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with
CC   a subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
```

Wed Feb 9 08:47:02 2000

CC   of the patient.
SQ   Sequence   10 AA;

Query Match   89.8%;   Score 44; DB 1; Length 10;
Best Local Similarity   90.0%; Pred. No. 0.0084;
Matches   9; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
          |:||||||||
Db   1 RENLRILLRY 10

RESULT   6
R95423
ID   R95423 standard; peptide; 10 AA.
AC   R95423;
DT   12-NOV-1996 (first entry)
DE   HLA-B2705.75-84
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 11; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702.75-84(D). These sequences can be used to isolate the protein
CC   p74 from a T-cell lysate. p74 is a T-cell surface membrane protein
CC   associated with T-cell activation in mammalian T-cells, and is also
CC   immunologically cross reactive with the heat shock protein Hsc70. p74 is
CC   found in a limited number of cell types, but is particularly expressed on
CC   B and T cells. p74 can be isolated by lysis of a suitable cell with an
CC   amphoteric detergent, and then passed through an affinity column
CC   containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence   10 AA.

Query Match   89.8%;   Score 44; DB 1; Length 10;
Best Local Similarity   90.0%; Pred. No. 0.0084;
Matches   9; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
          ||||| ||||
Db   1 REDLRTLLRY 10

RESULT   7
R95425
ID   R95425 standard; peptide; 10 AA.
AC   R95425;
DT   12-NOV-1996 (first entry)
DE   HLA-B2702.75-84(D).
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   misc_difference   3
FT   /note= "N3D mutation"
PN   WO9513288-A1. .
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 11; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702.75-84(D). These sequences can be used to isolate the protein
CC   p74 from a T-cell lysate. p74 is a T-cell surface membrane protein
CC   associated with T-cell activation in mammalian T-cells, and is also
CC   immunologically cross reactive with the heat shock protein Hsc70. p74 is
CC   found in a limited number of cell types, but is particularly expressed on
CC   B and T cells. p74 can be isolated by lysis of a suitable cell with an
CC   amphoteric detergent, and then passed through an affinity column
CC   containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence   10 AA.

Query Match   89.8%;   Score 44; DB 1; Length 10;
Best Local Similarity   90.0%; Pred. No. 0.0084;
Matches   9; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
          |||||| |||
Db   1 REDLRIALRY 10

RESULT   8
W07513
ID   W07513 standard; peptide; 10 AA.
AC   W07513;
DT   04-AUG-1997 (first entry)
DE   T-cell modulating peptide #2.
KW   T-cell modulator; autoimmune disease; tissue destruction; alpha1-domain;
KW   mammal; major histocompatability complex; MHC class I; antigen; perforin;
KW   insulin-dependent diabetes mellitus; multiple sclerosis; inflammation;
KW   rheumatoid arthritis; psoriasis; pemphigus vulgaris; Sjogren's disease;
KW   thyroid disease; Hashimoto's thyroiditis; myasthenia gravis; granzyme;
KW   autologous target cell; cytokine release; T cell activation; therapy.
OS   Synthetic.
PN   WO9635443-A1.
PD   14-NOV-1996.
PF   05-APR-1996; U04710.
PR   12-MAY-1995; US-440504.
PA   (SANG-) SANGSTAT MEDICAL CORP.
PI   Below R;
DR   WPI; 96-518410/51.
PT   Treatment of auto-immune disease by admin. of peptide(s) corresp. to
PT   major histocompatibility complex antigens - esp. for delaying onset
PT   of clinical symptoms of insulin dependent diabetes by modulating T
PT   cell mediated attack on target cells

PS Claim 7; Page 20; 24pp; English.
CC W07512-W07518 represent T-cell modulating peptides that can be used in
CC the method of the invention. These sequences are based on a portion of
CC the generic peptide corresponding to residues 70-91 of the alpha1-domain
CC of the major histocompatibility complex (MHC) class I antigen (see
CC W07510). The method is for affecting the course of an autoimmune disease
CC involving T-cell mediated destruction of tissue in mammals. These
CC peptides are used especially to treat insulin-dependent diabetes
CC mellitus, preferably being administered during the pre-clinical stage to
CC delay onset of the disease. Other diseases that can be treated are
CC multiple sclerosis, rheumatoid arthritis, psoriasis, pemphigus vulgaris,
CC Sjogren's disease, thyroid disease, Hashimoto's thyroiditis, myasthenia
CC gravis, etc. The peptides modulate T-cell mediated attack on autologous
CC target cells, and may also reduce inflammation, swelling, and release of
CC cytokines, perforins, granzymes etc. associated with T cell activation.
SQ Sequence 10 AA;

Query Match 89.8%; Score 44; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0084;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 REDLRILLRY 10
       ||:||| |||
Db 1 REDLRIALRY 10

RESULT 9
W47267
ID W47267 standard; peptide; 10 AA.
AC W47267;
DT 22-MAY-1998 (first entry)
DE Immunomodulatory peptide.
KW Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW transplant rejection; treatment; autoimmune disease.
OS Homo sapiens.
OS Synthetic.
FH Key   Location/Qualifiers
FT Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI; 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10; Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched
CC donor, i.e. to inhibit transplant rejection. It can also be used in
CC the treatment of autoimmune diseases.
CC Peptides using the D-form amino acids are more effective
CC immunomodulators than their diastereomers or enantiomers.
SQ Sequence 10 AA;

Query Match 89.8%; Score 44; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0084;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 REDLRILLRY 10
       ||:||||||
Db 1 RENLRILLRY 10

RESULT 10

W47269
ID W47269 standard; peptide; 10 AA.
AC W47269;
DT 22-MAY-1998 (first entry)
DE Immunomodulatory peptide.
KW Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW transplant rejection; treatment; autoimmune disease.
OS Homo sapiens.
OS Synthetic.
FH Key   Location/Qualifiers
FT Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI; 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10; Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched
CC donor, i.e. to inhibit transplant rejection. It can also be used in
CC the treatment of autoimmune diseases.
CC Peptides using the D-form amino acids are more effective
CC immunomodulators than their diastereomers or enantiomers.
SQ Sequence 10 AA;

Query Match 89.8%; Score 44; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0084;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 REDLRILLRY 10
       ||||| |||
Db 1 REDLRIALRY 10

RESULT 11
W33785
ID W33785 standard; peptide; 10 AA.
AC W33785;
DT 19-JUN-1998 (first entry)
DE Peptide B2705.75-84 tested for immunomodulating activity.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN WO9744351-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI; 98-086530/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or
PT treating autoimmune diseases
PS Example 1; Page 19; 41pp; English.
CC Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC activity. A peptide-type compound or variant is claimed which has
CC immunomodulating activity, including the N-terminal acylated and/or
CC C-terminal amidated or esterified forms of up to 60 amino acids, where
CC the peptide-type compound comprises the formula; A-B, where A, B -
CC (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R)); aa76 = E or V; aa77 -
CC D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or

```
CC    small amino acid: aa82 = R or L; aa83 = G or R; and aa represents amino
CC    acid. The sequence in the brackets may optionally be absent or truncated
CC    at any peptide type bond within the brackets. The compounds comprise
CC    amino acid sequences related to a Class I HLA-B alphal domain (positions
CC    79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC    undesirably attacking cells in a host or in vitro. They can also be
CC    used in combination with antigenic peptides or proteins of interest to
CC    activate CTLs. They can also inhibit the proliferation of T cells in
CC    response to anti-CD3. The peptide can be used for preventing rejection
CC    of transplants or for treating autoimmune diseases, e.g. diabetes,
CC    rheumatoid arthritis and lupus erythematosis. The products can also be
CC    used for detection and diagnosis.
SQ    Sequence   10 AA;

Query Match             89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity   90.0%;  Pred. No. 0.0084;
Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 REDLRILLRY 10
          ||||| ||||
Db      1 REDLRTLLRY 10

RESULT 12
W33787
ID   W33787 standard; peptide; 10 AA.
AC   W33787;
DT   19-JUN-1998  (first entry)
DE   Peptide B2702.75-84D77 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alphal domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   10 AA;

Query Match             89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity   90.0%;  Pred. No. 0.0084;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy      1 REDLRILLRY 10
          ||||| ||||
Db      1 REDLRTLLRY 10

RESULT 13
W33789
ID   W33789 standard; peptide; 10 AA.
AC   W33789;
DT   19-JUN-1998  (first entry)
DE   Peptide B2702.75-84L81 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alphal domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   10 AA;

Query Match             89.8%;  Score 44;  DB 1;  Length 10;
Best Local Similarity   90.0%;  Pred. No. 0.0084;
Matches    9;  Conservative    .1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 REDLRILLRY 10
          ||:||||||
Db      1 RENLRILLRY 10

RESULT 14
R71442
ID   R71442 standard; peptide; 17 AA.
AC   R71442;
DT   12-OCT-1995  (first entry)
DE   Human HLA-B27-(62-85) antigen derived peptide.
KW   Human HLA-B27-(62-85) antigen derived peptide; cell receptor;
KW   interaction modulation; arthritis; neoplasias; lupus erythematosus.
OS   Homo sapiens.
PN   WO9505189-A.
PD   23-FEB-1995.
PF   12-AUG-1994; U09189.
PR   12-AUG-1993; US-105416.
PA   (REGC ) UNIV CALIFORNIA.
PI   Goldstein A, Goodenow RS,  Olsson L;
DR   WPI; 95-098577/13.

Qy      1 REDLRILLRY 10
```

```
PT   Regulating cell surface receptor response - by modulating
PT   interaction between MHC class I antigen and the cell surface
PT   receptor
PS   Example 4; Page 45; 103pp; English.
CC   R71439-R71443 are human major histocompatibility complex class 1
CC   (MHC 1) alpha 1 domain and HLA derived peptides and fusion peptides.
CC   They were used to modulate interactions between MHC 1/HLA and cell
CC   surface receptors. Via competitive inhibition the peptides diminish
CC   the receptors response, this feature may be useful for the treatment
CC   of neoplasias, lupus erythematosus and arthritis.
SQ   Sequence    17 AA;

Query Match             89.8%;  Score 44;  DB 1;  Length 17;
Best Local Similarity   90.0%;  Pred. No. 0.015;
Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     1 REDLRILLRY 10
         ||||| ||||
Db     7 REDLRTLLRY 16


RESULT  15
R71443
ID   R71443 standard; peptide; 17 AA.
AC   R71443;
DT   12-OCT-1995  (first entry)
DE   Human [Phe74]-HLA-B27-(62-85) antigen derived peptide.
KW   Human [Phe74]-HLA-B27-(62-85) antigen derived peptide; cell receptor;
KW   interaction modulation; arthritis; neoplasias; lupus erythematosus.
OS   Homo sapiens.
PN   WO9505189-A.
PD   23-FEB-1995.
PF   12-AUG-1994; U09189.
PR   12-AUG-1993; US-105416.
PA   (REGC ) UNIV CALIFORNIA.
PI   Goldstein A, Goodenow RS,  Olsson L;
DR   WPI: 95-098577/13.
PT   Regulating cell surface receptor response - by modulating
PT   interaction between MHC class I antigen and the cell surface
PT   receptor
PS   Example 4; Page 45; 103pp; English.
CC   R71439-R71443 are human major histocompatibility complex class 1
CC   (MHC 1) alpha 1 domain and HLA derived peptides and fusion peptides.
CC   They were used to modulate interactions between MHC 1/HLA and cell
CC   surface receptors. Via competitive inhibition the peptides diminish
CC   the receptors response, this feature may be useful for the treatment
CC   of neoplasias, lupus erythematosus and arthritis.
SQ   Sequence    17 AA;

Query Match             89.8%;  Score 44;  DB 1;  Length 17;
Best Local Similarity   90.0%;  Pred. No. 0.015;
Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     1 REDLRILLRY 10
         ||||| ||||
Db     7 REDLRTLLRY 16


Search completed: February  8, 2000, 01:29:38
Job time: 1750 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February 7, 2000, 11:54:25 ; Search time 117.7 Seconds
                                          (without alignments)
                                          4.008 Million cell updates/sec

Title:        US-08-653-294-14
Perfect score: 49
Sequence:     1 REDLRILLRY 10

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched:     142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters:      142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :    PIR_62:*
              1:  pir1:*
              2:  pir2:*
              3:  pir3:*
              4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	338	2	I56116	MHC HLA-B27-HS - h
2	44	89.8	362	1	HLHUB2	MHC class I histoc
3	44	89.8	362	2	C35997	MHC class I histoc
4	44	89.8	362	2	I68724	MHC class I histoc
5	44	89.8	362	2	I37485	human lymphocyte a
6	44	89.8	362	2	I54289	MHC HLA-B27d - hum
7	39	79.6	137	2	I80174	class I histocompa
8	39	79.6	273	2	I38509	MHC class I histoc
9	39	79.6	354	2	I54463	MHC HLA-B38 chain
10	39	79.6	354	2	I59308	class I histocompa
11	39	79.6	354	2	I80168	class I histocompa
12	39	79.6	354	2	I80167	class I histocompa
13	39	79.6	355	2	I80169	class I histocompa
14	39	79.6	355	2	I80171	class I histocompa
15	39	79.6	359	1	HLHU12	MHC class I histoc
16	39	79.6	362	1	HLHUB8	MHC class I histoc
17	39	79.6	362	2	B30345	MHC class I histoc
18	39	79.6	362	2	JH0541	class I histocompa
19	39	79.6	362	2	JH0539	class I histocompa
20	39	79.6	362	2	JH0540	class I histocompa
21	39	79.6	362	2	A45834	MHC class I histoc
22	39	79.6	362	2	I84486	transmembrane glyc
23	39	79.6	362	2	I62045	gene HLA B-1517 pr
24	39	79.6	362	2	I84490	lymphocyte antigen
25	39	79.6	362	2	I37521	HLA-Bw57.2 antigen
26	39	79.6	362	2	A30345	MHC class I histoc
27	39	79.6	362	2	I59633	MHC HLA-B transmem
28	39	79.6	362	2	S24434	class I histocompa
29	39	79.6	362	2	I37120	MHC class I histoc
30	39	79.6	363	2	S07113	class I histocompa
31	39	79.6	363	2	S03537	class I histocompa
32	39	79.6	364	2	D35997	MHC class I histoc
33	39	79.6	365	2	S77963	MHC class I histoc
34	39	79.6	365	2	I54416	HLA-AW24 protein -
35	39	79.6	365	2	I54493	MHC class I histoc
36	38	77.6	274	2	HLHU32	MHC class I histoc
37	38	77.6	362	1	I37515	MHC class I histoc
38	38	77.6	364	2	A35997	MHC class I histoc
39	37	75.5	328	2	I54414	MHC H-2K transplan
40	37	75.5	362	2	I71998	MHC H-2D-k protein
41	37	75.5	368	2	I68705	MHC H-2K-w28 prote
42	34	69.4	339	2	T15113	hypothetical prote
43	34	69.4	355	2	I37516	HLA-B alpha-chain
44	34	69.4	358	1	ADMU	fructose-bisphosph
45	34	69.4	358	2	T05052	fructose-bisphosph

ALIGNMENTS

```
RESULT  1
I56116
MHC HLA-B27-HS - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C;Accession: I56116
R;Choo, S.Y.;Fan, L.A.;Hansen, J.A.
J. Immunol. 147, 174-180, 1991
A;Title: A novel HLA-B27 allele maps B27 allospecificity to the region around positio
A;Reference number: I56116; MUID:91268545
A;Accession: I56116
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-338 <RES>
A;Cross-references: GB:M62852; NID:g187760; PIDN:AAA59647.1; PID:g187761
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology


Query Match             89.8%;  Best Local Similarity   90.0%;
Matches   9;  Conservative   0;  Mismatches   0;
                                                 Score 44;  DB 2;  Length 338;
                                                 Pred. No. 0.2;
                                                 Indels   0;  Gaps   0;

Qy        1 REDLRILLRY 10
            ||||| ||||
Db       75 REDLRTLLRY 84


RESULT  2
HLHUB2
MHC class I histocompatibility antigen HLA-B27 alpha chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 28-Apr-1995 #text_change 22-Jun-1999
C;Accession: S07441; A25092; B25092; A94087; S34180; S44942; A90493; B24741; I55965;
R;Weiss, E.H.; Kuon, W.; Doerner, C.; Lang, M.; Riethmueller, G.
Immunobiology 170, 367-380, 1985
A;Title: Organization, sequence and expression of the HLA-B27 gene: a molecular appro
A;Reference number: S07441; MUID:86138405
A;Accession: S07441
A;Molecule type: DNA
A;Residues: 1-362 <WEI>
A;Cross-references: EMBL:X03945
A;Note: the authors translated the codon GAC for residue 61 as Ala and the codon CAG
R;Seemann, G.H.A.; Rein, R.S.; Brown, C.S.; Ploegh, H.L.
EMBO J. 5, 547-552, 1986
A;Title: Gene conversion-like mechanisms may generate polymorphism in human class I g
A;Reference number: A91061; MUID:86220133
A;Accession: A25092
A;Molecule type: DNA
A;Residues: 1-362 <SEE>
A;Cross-references: GB:X03665; NID:g32250; PIDN:CAA27302.1; PID:g871297.
A;Note: this allele is designated B*27051 (formerly 27W)
A;Accession: B25092
```

A:Molecule type: DNA
A:Residues: 1-100,'N',102-103,'IA',106-362 <SE2>
A:Cross-references: GB:X03664; NID:g32236; PIDN:CAA27301.1; PID:g871296
A:Note: this allele is designated B*2702 (formerly 27K)
R:Szoets, H.; Riethmueller, G.; Weiss, E.; Meo, T.
Proc. Natl. Acad. Sci. U.S.A. 83, 1428-1432, 1986
A:Title: Complete sequence of HLA-B27 cDNA identified through the characterization of st
A:Reference number: A94087. MUID:86149317
A:Accession: A94087
A:Molecule type: mRNA
A:Residues: 25-205,'V',207-362 <SZO>
A:Cross-references: GB:M12678
A:Note: this allele is designated B*27052 (formerly 27W)
R:Vilches, C.
submitted to the EMBL Data Library, June 1993
A:Reference number: S34180
A:Accession: S34180
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100,'S',102-137,'D',139,'Y',141-175,'E',177-362 <VIL>
A:Cross-references: EMBL:X73578
A:Note: this allele is designated B*2706
R:D'Amato, M.; Sorrentino, R.
submitted to the EMBL Data Library, May 1994
A:Description: Identification of a novel HLA-B27 subtype by restriction analysis of a cy
A:Reference number: S44942
A:Accession: S44942
A:Molecule type: mRNA
A:Residues: 1-139,'H',141-362 <DAM>
A:Cross-references: EMBL:233453; NID:g486652; PIDN:CAA83876.1; PID:g486653
R:Ezquerra, A.; Bragado, R.; Vega, M.A.; Stronninger, J.L.; Woody, J.; Lopez de Castro, J
Biochemistry 24, 1733-1741, 1985
A:Title: Primary structure of papain-solubilized human histocompatibility antigen HLA-B2
A:Reference number: A90493; MUID:85226361
A:Accession: A90493
A:Molecule type: protein
A:Residues: 25-265,'E',267-295 <E2Q>
R:Vega, M.A.; Ezquerra, A.; Rojo, S.; Aparicio, P.; Bragado, R.; Lopez de Castro, J.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 7394-7398, 1985
A:Title: Structural analysis of an HLA-B27 functional variant: identification of residue
A:Reference number: A94070; MUID:86042671
A:Accession: B24741
A:Molecule type: protein
A:Residues: 86-100,'N',102-103,'IA',106-107;171-181 <VEG>
R:Coppin, H.L.; McDevitt, H.O.
J. Immunol. 137, 2168-2172, 1986
A:Title: Absence of polymorphism between HLA-B27 genomic exon sequences isolated from no
A:Reference number: I55965; MUID:87009855
A:Accession: I55965
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 25-298 <RES>
A:Cross-references: GB:M14013; NID:g187743; PIDN:AAA59643.1; PID:g187744
R:Blaszyk, R.; Weber, M.; Salama, A
submitted to the EMBL Data Library, January 1995
A:Reference number: S52291
A:Accession: S52291
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 116-192 <BLA>
A:Cross-references: EMBL:X83737
C:Comment: This allele for HLA-B correlates with the development of ankylosing spondylit
C:Genetics:
A:Gene: GDB:HLA-B
A:Cross-references: GDB:120048; OMIM:142830
A:Map position: 6p21.3-6p21.3
A:Introns: 25/1; 114/1; 207/1; 299/1; 338/1; 349/1
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: ankylosing spondylitis; duplication; glycoprotein; heterodimer; transmembran
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-362/Product: class I histocompatibility antigen HLA-B27 alpha chain #status predict
F:25-307/Domain: extracellular #status predicted <EXT>
F:25-114/Domain: alpha-1 <EX1>
F:115-206/Domain: alpha-2 <EX2>

F:220-285/Domain: immunoglobulin homology <IMM>
F:308-331/Domain: transmembrane #status predicted <TMM>
F:332-362/Domain: intracellular #status predicted <INT>
F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:125-188,227-283/Disulfide bonds: #status experimental

Query Match   89.8%;   Score 44; DB 1;   Length 362;
Best Local Similarity   90.0%;   Pred. No. 0.22;
Matches   9; Conservative   0; Mismatches   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
         ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   3
C35997
MHC class I histocompatibility antigen HLA-B37 alpha chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Nov-1990 #sequence_revision 13-Jan-1993 #text_change 23-Jul-1999
C:Accession: C35997
R:Ennis, P.D.; Zemmour, J.; Salter, R.D.; Parham, P.
Proc. Natl. Acad. Sci. U.S.A. 87, 2833-2837, 1990
A:Title: Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: freque
A:Reference number: A35997; MUID:90207291
A:Accession: C35997
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-362 <ENN>
A:Cross-references: GB:M32320; NID:g187792; PIDN:AAA36233.1; PID:g307224
C:Genetics:
A:Gene: GDB:HLA-B
A:Cross-references: GDB:120048; OMIM:142830
A:Map position: 6p21.3-6p21.3
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: transmembrane protein
F:220-285/Domain: immunoglobulin homology <IMM>

Query Match   89.8%;   Score 44; DB 2;   Length 362;
Best Local Similarity   90.0%;   Pred. No. 0.22;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
         ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   4
I68724
MHC class I histocompatibility antigen HLA-Bw47 precursor - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 23-Jul-1999
C:Accession: I68724
R:Zemmour, J.; Ennis, P.D.; Parham, P.; Dupont, B.
Immunogenetics 27, 281-287, 1988
A:Title: Comparison of the structure of HLA-Bw47 to HLA-B13 and its relationship to 2
A:Reference number: I54442; MUID:88152906
A:Accession: I68724
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-362 <ZEM>
A:Cross-references: GB:M19756; NID:g184171; PIDN:AAA52664.1; PID:g386776
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   89.8%;   Score 44; DB 2;   Length 362;
Best Local Similarity   90.0%;   Pred. No. 0.22;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
         ||||| ||||

Db   99 REDLRTLLRY 108

RESULT   5
I37485
human lymphocyte antigen HLA-B27 - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C;Accession: I37485
R;Del Porto, P.; D'Amato, M.; Fiorillo, M.T.; Tuosto, L.; Piccolella, E.; Sorrentino, R.
J. Immunol. 153, 3093-3100, 1994
A;Title: Identification of a novel HLA-B27 subtype by restriction analysis of a cytotoxi
A;Reference number: I37485; MUID:94375872
A;Accession: I37485
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-362 <RES>
A;Cross-references: EMBL:Z33453; NID:g486652; PIDN:CAA83876.1; PID:g486653
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   89.8%;   Score 44; DB 2;   Length 362;
Best Local Similarity   90.0%;   Pred. No. 0.22;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
         ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   6
I54289
MHC HLA-B27d - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C;Accession: I54289
R;Choo, S.Y.; St. John, T.; Orr, H.T.; Hansen, J.A.
Hum. Immunol. 21, 209-219, 1988
A;Title: Molecular analysis of the variant alloantigen HLA-B27d (HLA-B*2703) identifies
A;Reference number: I54289; MUID:88227491
A;Accession: I54289
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: GB:M54883; NID:g187663; PIDN:AAA59616.1; PID:g187664
C;Genetics:
A;Gene: GDB:HLA-B
A;Cross-references: GDB:120048; OMIM:142830
A;Map position: 6p21.3-6p21.3
A;Introns: 25/1; 115/1; 207/1; 299/1; 338/1; 349/1
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   89.8%;   Score 44; DB 2;   Length 362;
Best Local Similarity   90.0%;   Pred. No. 0.22;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
         ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   7
I80174
class I histocompatibility antigen - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C;Accession: I80174
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Watkin
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I80174
A;Accession: I80174

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-137 <RES>
A;Cross-references: EMBL:U05585; NID:g454787; PIDN:AAA50188.1; PID:g454788
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   79.6%;   Score 39; DB 2;   Length 137;
Best Local Similarity   80.0%;   Pred. No. 0.82;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
         ||:|| ||||
Db   40 RENLRTLLRY 49

RESULT   8
I38509
MHC class I histocompatibility antigen - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 23-Jul-1999
C;Accession: I38509
R;Cereb, N.; Choi, J.W.; Riu, K.Z.; Yang, S.Y.
Tissue Antigens 44, 271-273, 1994
A;Title: HLA-B*5105, a newly identified B51 IEF variant.
A;Reference number: I38509; MUID:95176331
A;Accession: I38509
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-273 <RES>
A;Cross-references: EMBL:U06697; NID:g469544; PIDN:AAA92997.1; PID:g469545
C;Genetics:
A;Gene: GDB:HLA-B
A;Cross-references: GDB:120048; OMIM:142830
A;Map position: 6p21.3-6p21.3
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   79.6%;   Score 39; DB 2;   Length 273;
Best Local Similarity   80.0%;   Pred. No. 1.7;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
         ||:|| ||||
Db   74 RENLRIALRY 83

RESULT   9
I54463
MHC HLA-B38 chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 23-Jul-1999
C;Accession: I54463
R;Mueller, C.A.; Engler-Blum, G.; Gekeler, V.; Steiert, I.; Weiss, E.; Schmidt, H.
Immunogenetics 30, 200-207, 1989
A;Title: Genetic and serological heterogeneity of the supertypic HLA-B locus specific
A;Reference number: I54463; MUID:89379286
A;Accession: I54463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-274 <RES>
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match   79.6%;   Score 39; DB 2;   Length 274;
Best Local Similarity   80.0%;   Pred. No. 1.7;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
         ||:|| ||||
Db   75 RENLRIALRY 84

```
RESULT 10
I59308
class I histocompatibility antigen - pygmy chimpanzee (fragment)
C;Species: Pan paniscus (pygmy chimpanzee, bonobo)
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 23-Jul-1999
C;Accession: I59308
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Watkin
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I59308; MUID:94286544
A;Accession: I59308
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-354 <RES>
A;Cross-references: EMBL:U05575; NID:g454767; PIDN:AAA50178.1; PID:g454768
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match      79.6%;  Score 39;  DB 2;  Length 354;
Best Local Similarity  80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1 REDLRILLRY 10
        ||:|| |||
Db  91 RENLRIALRY 100

RESULT 11
I80168
class I histocompatibility antigen - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C;Accession: I80168
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Watkin
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I59308; MUID:94286544
A;Accession: I80168
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-354 <RES>
A;Cross-references: EMBL:U05579; NID:g454775; PIDN:AAA50182.1; PID:g454776
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match      79.6%;  Score 39;  DB 2;  Length 354;
Best Local Similarity  80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1 REDLRILLRY 10
        ||:|| |||
Db  91 RENLRIALRY 100

RESULT 12
I80167
class I histocompatibility antigen - pygmy chimpanzee (fragment)
C;Species: Pan paniscus (pygmy chimpanzee, bonobo)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C;Accession: I80167
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Watkin
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I59308; MUID:94286544
A;Accession: I80167
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-354 <RES>
A;Cross-references: EMBL:U05578; NID:g454773; PIDN:AAA50181.1; PID:g454774
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match      79.6%;  Score 39;  DB 2;  Length 354;
Best Local Similarity  80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1 REDLRILLRY 10
        ||:|| |||
Db  91 RENLRIALRY 100

RESULT 13
I80169
class I histocompatibility antigen - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C;Accession: I80169
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I59308; MUID:94286544
A;Accession: I80169
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: EMBL:U05580; NID:g454777; PIDN:AAA50183.1; PID:g454778
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match      79.6%;  Score 39;  DB 2;  Length 355;
Best Local Similarity  80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1 REDLRILLRY 10
        ||:|| |||
Db  91 RENLRIALRY 100

RESULT 14
I80171
class I histocompatibility antigen - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 23-Jul-1999
C;Accession: I80171
R;McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes, A.L.; Bontrop, R.E.; Wat
Proc. Natl. Acad. Sci. U.S.A. 91, 5893-5897, 1994
A;Title: A uniquely high level of recombination at the HLA-B locus.
A;Reference number: I59308; MUID:94286544
A;Accession: I80171
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: EMBL:U05582; NID:g454781; PIDN:AAA50185.1; PID:g454782
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match      79.6%;  Score 39;  DB 2;  Length 355;
Best Local Similarity  80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1 REDLRILLRY 10
        ||:|| |||
Db  91 RENLRIALRY 100

RESULT 15
HLHU12
MHC class I histocompatibility antigen HLA alpha chain precursor (clone pHLA 12.4) -
C;Species: Homo sapiens (man)
C;Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 22-Jun-1999
C;Accession: A02189
R;Malissen, M.; Malissen, B.; Jordan, B.R.
Proc. Natl. Acad. Sci. U.S.A. 79, 893-897, 1982
A;Title: Exon/intron organization and complete nucleotide sequence of an HLA gene.
A;Reference number: A02189; MUID:82151002
```

```
A;Accession: A02189
A;Molecule type: DNA
A;Residues: 1-359 <MAL>
A;Cross-references: GB:J00191; GB:V00526; NID:g187600; PIDN:AAA36218.1; PID:g386873
C;Comment: The seven exons correspond approximately to the domain structure of this chai
C;Genetics:
A;Map position: 6p21.3
A;Introns: 22/1; 112/1; 204/1; 296/1; 335/1; 346/1
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology
C;Keywords: duplication; glycoprotein; heterodimer; transmembrane protein; transplantati
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-359/Product: class I histocompatibility antigen HLA alpha chain #status predicted <
F;22-304/Domain: extracellular #status predicted <EXT>
F;22-111/Domain: alpha-1 <EX1>
F;112-203/Domain: alpha-2 <EX2>
F;217-282/Domain: immunoglobulin homology <IMM>
F;305-329/Domain: transmembrane #status predicted <TMM>
F;335-359/Domain: intracellular #status predicted <INT>
F;107/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;224-280/Disulfide bonds: #status predicted


Query Match             79.6%;  Score 39;  DB 1;  Length 359;
Best Local Similarity   80.0%;  Pred. No. 2.2;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 REDLRILLRY 10
          ||:||| |||
Db   96 RENLRIALRY 105


Search completed: February  7, 2000, 11:54:25
Job time: 24335 sec
```

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

```
          GenCore version 4.5
    Copyright (C) 1993 - 2000 Compugen Ltd.
```

Run on: February 8, 2000, 00:59:53 ; Search time 63.71 Seconds
                                            (without alignments)
                                            4.688 Million cell updates/sec

Title: US-08-653-294-14
Perfect score: 49
Sequence: 1 REDLRILLRY 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters: 82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	338	1	1B20_HUMAN	P30467 homo sapien
2	44	89.8	361	1	1B14_HUMAN	P03989 homo sapien
3	44	89.8	362	1	1B16_HUMAN	P19373 homo sapien
4	44	89.8	362	1	1B18_HUMAN	P10318 homo sapien
5	44	89.8	362	1	1B29_HUMAN	P18463 homo sapien
6	44	89.8	362	1	1B45_HUMAN	P30485 homo sapien
7	39	79.6	359	1	1B01_PANTR	P13750 pan troglod
8	39	79.6	362	1	1B01_GORGO	P30379 gorilla gor
9	39	79.6	362	1	1B02_GORGO	P30380 gorilla gor
10	39	79.6	362	1	1B03_GORGO	P30381 gorilla gor
11	39	79.6	362	1	1B15_HUMAN	P10317 homo sapien
12	39	79.6	362	1	1B47_HUMAN	P30487 homo sapien
13	39	79.6	362	1	1B49_HUMAN	P18464 homo sapien
14	39	79.6	362	1	1B52_HUMAN	P30489 homo sapien
15	39	79.6	362	1	1B53_HUMAN	P30490 homo sapien
16	39	79.6	362	1	1B54_HUMAN	P30491 homo sapien
17	39	79.6	362	1	1B60_HUMAN	P18465 homo sapien
18	39	79.6	362	1	1B61_HUMAN	P30497 homo sapien
19	39	79.6	362	1	1B62_HUMAN	P10319 homo sapien
20	39	79.6	362	1	HLAH_HUMAN	P01893 homo sapien
21	39	79.6	365	1	1A23_HUMAN	P30447 homo sapien
22	39	79.6	365	1	1A24_HUMAN	P05534 homo sapien
23	38	77.6	362	1	1B19_HUMAN	Q08136 homo sapien
24	38	77.6	365	1	1B25_HUMAN	P18462 homo sapien
25	38	77.6	365	1	1A32_HUMAN	P10314 homo sapien
26	37	75.5	328	1	HA1Q_MOUSE	P14428 mus musculu
27	37	75.5	362	1	HA13_MOUSE	P14426 mus musculu
28	37	75.5	362	1	HA1W_MOUSE	P03991 mus musculu
29	34	69.4	358	1	ALF_ARATH	P22197 arabidopsis
30	34	69.4	362	1	1B05_HUMAN	P30461 homo sapien
31	34	69.4	362	1	1B41_HUMAN	P30481 homo sapien
32	34	69.4	362	1	1B42_HUMAN	P30482 homo sapien
33	33	67.3	365	1	1A04_GORGO	P30378 gorilla gor
34	33	67.3	403	1	FLGE_HELMU	P50611 helicobacte
35	33	67.3	501	1	YM05_YEAST	Q04472 saccharomyc
36	33	67.3	599	1	SYD_SYNY3	P73851 synechocyst
37	33	67.3	633	1	KS63_MOUSE	P18654 mus musculu
38	33	67.3	724	1	KS61_MOUSE	P18653 mus musculu
39	33	67.3	733	1	KS62_HUMAN	Q15349 homo sapien
40	33	67.3	735	1	KS61_HUMAN	Q15418 homo sapien
41	33	67.3	740	1	KS63_HUMAN	P51812 homo sapien
42	33	67.3	752	1	KS6A_CHICK	P18652 gallus gall
43	32	65.3	157	1	Y188_RICPR	Q9zdx6 rickettsia
44	32	65.3	173	1	PHB3_FREDI	P14877 fremyella d
45	32	65.3	252	1	YKC6_CAEEL	P41997 caenorhabdi

ALIGNMENTS

RESULT 1
1B20_HUMAN
ID   1B20_HUMAN   STANDARD;   PRT;   338 AA.
AC   P30467;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2707 ALPHA CHAIN
DE   (B27-HS).
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91268545.
RA   CHOO Y.S., FAN L.A., HANSEN J.A.;
RT   "A novel HLA-B27 allele maps B27 allospecificity to the region around
RT   position 70 in the alpha 1 domain.";
RL   J. Immunol. 147:174-180(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M62852; AAA59647.1; -.
DR   HSSP; P03989; 1HSA.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein.
FT   DOMAIN   1   90   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   91   182   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   183   274   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   275   284   CONNECTING PEPTIDE.
FT   TRANSMEM   285   308
FT   DOMAIN   309   338   CYTOPLASMIC TAIL.
FT   CARBOHYD   86   86   BY SIMILARITY.
FT   DISULFID   101   164   BY SIMILARITY.
FT   DISULFID   203   259   BY SIMILARITY.
SQ   SEQUENCE   338 AA; 37804 MW; 33FB8134 CRC32;

Query Match   89.8%; Score 44; DB 1; Length 338;
Best Local Similarity   90.0%; Pred. No. 0.11;
Matches   9; Conservative   0; Mismatches   1; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10

```
                        |||| ||||
Db       75 REDLRTLLRY 84
```

RESULT 2
ID   1B14_HUMAN   STANDARD;   PRT;   361 AA.
AC   P03989;
DT   23-OCT-1986 (Rel. 02, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86138405.
RA   WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;
RT   "Organization, sequence and expression of the HLA-B27 gene: a
RT   molecular approach to analyze HLA and disease association.";
RL   Immunobiology 170:367-380(1985).
RN   [2]
RP   SEQUENCE OF 25-361 FROM N.A.
RX   MEDLINE; 86149317.
RA   SZOETS H., RIETHMUELLER G., WEISS E., MEO T.;
RT   "Complete sequence of HLA-B27 cDNA identified through the
RT   characterization of structural markers unique to the HLA-A, -B, and
RT   -C allelic series.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1428-1432(1986).
RN   [3]
RP   SEQUENCE OF 25-295.
RX   MEDLINE; 85226361.
RA   EZQUERRA A., BRAGADO R., VEGA M.A., STROMINGER J.L., WOODY J.,
RA   LOPEZ DE CASTRO J.A.;
RT   "Primary structure of papain-solubilized human histocompatibility
RT   antigen HLA-B27.";
RL   Biochemistry 24:1733-1741(1985).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-300.
RX   MEDLINE; 92405152.
RA   MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;
RT   "The three-dimensional structure of HLA-B27 at 2.1-A resolution
RT   suggests a general mechanism for tight peptide binding to MHC.";
RL   Cell 70:1035-1048(1992).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE; 92018187.
RA   MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;
RT   "The structure of HLA-B27 reveals nonamer self-peptides bound in an
RT   extended conformation.";
RL   Nature 353:321-325(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   -!- DISEASE: THIS PROTEIN CORRELATES WITH THE DEVELOPMENT OF
CC   ANKYLOSING SPONDYLITIS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X03945; CAA27578.1; ALT_TERM.
DR   PIR; A25128; HLHUB2.
DR   PIR; S07441; S07441.
DR   PDB; 1HSA; 15-OCT-92.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL   1   24
FT   CHAIN   25   361   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT   B-27 ALPHA CHAIN.
FT   DOMAIN   25   114   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   115   206   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   207   298   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   299   308   CONNECTING PEPTIDE.
FT   TRANSMEM   309   332
FT   DOMAIN   333   361   CYTOPLASMIC TAIL.
FT   CARBOHYD   110   110
FT   DISULFID   125   188
FT   DISULFID   227   283
FT   CONFLICT   206   206   A -> V (IN REF. 2).
FT   CONFLICT   266   266   Q -> E (IN REF. 3).
FT   STRAND   27   38
FT   TURN   39   41
FT   STRAND   42   52
FT   TURN   53   54
FT   STRAND   55   61
FT   TURN   62   63
FT   STRAND   70   71
FT   HELIX   74   76
FT   TURN   77   78
FT   HELIX   81   108
FT   TURN   109   110
FT   TURN   113   114
FT   STRAND   118   127
FT   TURN   129   130
FT   STRAND   133   142
FT   TURN   143   144
FT   STRAND   145   150
FT   TURN   152   153
FT   STRAND   157   159
FT   HELIX   162   173
FT   TURN   174   175
FT   HELIX   176   185
FT   TURN   186   186
FT   HELIX   187   198
FT   HELIX   199   199
FT   TURN   200   203
FT   HELIX   204   204
FT   TURN   207   207
FT   TURN   210   217
FT   STRAND   222   233
FT   STRAND   238   243
FT   TURN   244   245
FT   STRAND   246   247
FT   HELIX   249   251
FT   STRAND   253   254
FT   STRAND   258   259
FT   STRAND   265   274
FT   TURN   275   276
FT   HELIX   278   280
FT   STRAND   281   286
FT   TURN   288   289
FT   STRAND   294   296
SQ   SEQUENCE   361 AA;   40464 MW; 802130D5 CRC32;

Query Match   89.8%;   Score 44; DB 1; Length 361;
Best Local Similarity   90.0%;   Pred. No. 0.12;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

```
Qy      1 REDLRILLRY 10
          |||| |||||
Db     99 REDLRTLLRY 108
```

RESULT   3

```
1B16_HUMAN
ID   1B16_HUMAN     STANDARD;      PRT;    362 AA.
AC   P19373;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2703 ALPHA CHAIN
DE   PRECURSOR (B-27D).
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88227491.
RA   CHOO S.Y., ST JOHN T., ORR H.T., HANSEN J.A.;
RT   "Molecular analysis of the variant alloantigen HLA-B27d (HLA-B*2703)
RT   identifies a unique single amino acid substitution.";
RL   Hum. Immunol. 21:209-219(1988)
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M54883; AAA59616.1; -.
DR   HSSP; P03989; 1HSA.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                B-27 B*2703 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    308       CONNECTING PEPTIDE.
FT   TRANSMEM    309    332
FT   DOMAIN      333    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE   362 AA;  40402 MW;  7261C3AB CRC32;

Query Match             89.8%;  Score 44;  DB 1;  Length 362;
Best Local Similarity   90.0%;  Pred. No. 0.12;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy      1 REDLRILLRY 10
          |||||  ||||
Db     99 REDLRTLLRY 108

RESULT   4
1B18_HUMAN
ID   1B18_HUMAN     STANDARD;      PRT;    362 AA.
AC   P10318;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2705 ALPHA CHAIN
DE   PRECURSOR (B-27W) (B27.1).
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86220133.
RA   SEEMANN G.H.A., REIN R.S., BROWN C.S., PLOEGH H.L.;
RT   "Gene conversion-like mechanisms may generate polymorphism in human
RT   class I genes.";
RL   EMBO J. 5:547-552(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86138405.
RA   WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;
RT   "Organization, sequence and expression of the HLA-B27 gene: a
RT   molecular approach to analyze HLA and disease associations.";
RL   Immunobiology 170:367-380(1985).
RN   [3]
RP   3D-STRUCTURE MODELING OF 115-206.
RX   MEDLINE; 95148615.
RA   ROGNAN D., SCAPOZZA L., FOLKERS G., DASER A.;
RT   "Rational design of nonnatural peptides as high-affinity ligands for
RT   the HLA-B*2705 human leukocyte antigen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:753-757(1995).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X03665; CAA27302.1; -.
DR   EMBL; X03666; CAA27302.1; JOINED.
DR   EMBL; M12967; AAA36221.1; -.
DR   PIR; A25092; HLHUBW.
DR   PDB; 1ROG; 30-SEP-94.
DR   PDB; 1ROH; 30-SEP-94.
DR   PDB; 1ROI; 30-SEP-94.
DR   PDB; 1ROJ; 30-SEP-94.
DR   PDB; 1ROK; 30-SEP-94.
DR   PDB; 1ROL; 30-SEP-94.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                B-27 B*2705 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    308       CONNECTING PEPTIDE.
FT   TRANSMEM    309    332
FT   DOMAIN      333    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE   362 AA;  40428 MW;  73243566 CRC32;

Query Match             89.8%;  Score 44;  DB 1;  Length 362;
Best Local Similarity   90.0%;  Pred. No. 0.12;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy      1 REDLRILLRY 10
          ||||| |||||
```

```
Db     99 REDLRTLLRY 108

RESULT 5
1B29_HUMAN
ID   1B29_HUMAN     STANDARD;      PRT;   362 AA.
AC   P18463;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-37 B*3701 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90207291.
RA   ENNIS P.D., ZEMMOUR J., SALTER R.D., PARHAM P.;
RT   "Rapid cloning of HLA-A,B cDNA by using the polymerase chain
RT   reaction: frequency and nature of errors produced in amplification.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2833-2837(1990).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M32320; AAA36233.1; -.
DR   PIR; C35997; C35997.
DR   HSSP; P03989; 1HSA.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                B-37 B*3701 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    308       CONNECTING PEPTIDE.
FT   TRANSMEM    309    332
FT   DOMAIN      333    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE   362 AA;  40456 MW;  8C567F8E CRC32;

Query Match             89.8%;  Score 44;  DB 1;  Length 362;
Best Local Similarity   90.0%;  Pred. No. 0.12;
Matches  . 9;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy      1 REDLRILLRY 10
          ||||| ||||
Db     99 REDLRTLLRY 108

RESULT 6
1B45_HUMAN
ID   1B45_HUMAN     STANDARD;      PRT;   362 AA.
AC   P30485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-47 B*4701 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88152906.
RA   ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.;
RT   "Comparison of the structure of HLA-Bw47 to HLA-B13 and its
RT   relationship to 21-hydroxylase deficiency.";
RL   Immunogenetics 27:281-287(1988).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M19756; AAA2664.1; -.
DR   HSSP; P03989; 1HSA.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                BW-47 B*4701 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    309       CONNECTING PEPTIDE.
FT   TRANSMEM    310    333
FT   DOMAIN      334    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE   362 AA;  40571 MW;  67734C1E CRC32;

Query Match             89.8%;  Score 44;  DB 1;  Length 362;
Best Local Similarity   90.0%;  Pred. No. 0.12;
Matches   9;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy      1 REDLRILLRY 10
          ||||| ||||
Db     99 REDLRTLLRY 108

RESULT 7
1B01_PANTR
ID   1B01_PANTR     STANDARD;      PRT;     359 AA.
AC   P13750;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR
DE   (FRAGMENT).
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN   [1]
RP   SEQUENCE FROM N.A.
```

RX   MEDLINE; 89030641.
RA   MAYER W.E., JONKER M., KLEIN D., IVANYI P., VAN SEVENTER G.,
RA   KLEIN J.;
RT   "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for
RT   trans-species mode of evolution.";
RL   EMBO J. 7:2765-2774(1988).
RN   [2]
RP   REVISIONS.
RA   MAYER W.;
RL   Submitted (FEB-1989) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X13115; CAA31507.1; -.
DR   PIR; S03537; S03537.
DR   HSSP; P03989; 1HSA.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   NON_TER   1   1
FT   SIGNAL   <1   20
FT   CHAIN   21   359   CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT   B-1 ALPHA CHAIN.
FT   DOMAIN   21   110   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   111   202   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   203   294   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   295   305   CONNECTING PEPTIDE.
FT   TRANSMEM   306   329
FT   DOMAIN   330   359   CYTOPLASMIC TAIL.
FT   DISULFID   121   184   BY SIMILARITY.
FT   DISULFID   223   279   BY SIMILARITY.
FT   CARBOHYD   106   106   BY SIMILARITY.
SQ   SEQUENCE   359 AA; 40173 MW; 5395FFC9 CRC32;

Query Match   79.6%; Score 39; DB 1; Length 359;
Best Local Similarity   80.0%; Pred. No. 1.2;
Matches   8; Conservative   8; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
           ||:||| |||
Db   95 RENLRIALRY 104

RESULT 8
1B01_GORGO
ID   1B01_GORGO   STANDARD;   PRT;   362 AA.
AC   P30379;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0101 ALPHA CHAIN PRECURSOR.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT   "Gorilla class I major histocompatibility complex alleles: comparison
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).

CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X60255; CAA42807.1; -.
DR   PIR; JH0539; JH0539.
DR   HSSP; P03989; 1HSA.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL   1   24   BY SIMILARITY.
FT   CHAIN   25   362   CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT   GOGO-B0101 ALPHA CHAIN.
FT   DOMAIN   25   114   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   115   206   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   207   298   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   299   308   CONNECTING PEPTIDE.
FT   TRANSMEM   309   332
FT   DOMAIN   333   362   CYTOPLASMIC TAIL.
FT   DISULFID   125   188   BY SIMILARITY.
FT   DISULFID   227   283   BY SIMILARITY.
FT   CARBOHYD   110   110   BY SIMILARITY.
SQ   SEQUENCE   362 AA; 40170 MW; 2E33E2B8 CRC32;

Query Match   79.6%; Score 39; DB 1; Length 362;
Best Local Similarity   80.0%; Pred. No. 1.2;
Matches   8; Conservative   8; Mismatches   1; Indels   0; Gaps   0;

Qy   1 REDLRILLRY 10
           ||:||| |||
Db   99 RENLRIALRY 108

RESULT 9
1B02_GORGO
ID   1B02_GORGO   STANDARD;   PRT;   362 AA.
AC   P30380;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0102 ALPHA CHAIN PRECURSOR.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT   "Gorilla class I major histocompatibility complex alleles: comparison
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC    or send an email to license@isb-sib.ch).
CC    ------------------------------------------------------------
DR    EMBL; X60693; CAA43101.1; -.
DR    PIR; JH0540; JH0540.
DR    HSSP; P03989; 1HSA.
DR    PROSITE; PS00290; IG_MHC; 1.
DR    PFAM; PF00047; ig; 1.
DR    PFAM; PF00129; MHC_I; 1.
KW    MHC I; Transmembrane; Glycoprotein; Signal.
FT    SIGNAL       1     24            BY SIMILARITY.
FT    CHAIN       25    362            CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                     GOGO-B0102 ALPHA CHAIN.
FT    DOMAIN      25    114            EXTRACELLULAR ALPHA-1.
FT    DOMAIN     115    206            EXTRACELLULAR ALPHA-2.
FT    DOMAIN     207    298            EXTRACELLULAR ALPHA-3.
FT    DOMAIN     299    308            CONNECTING PEPTIDE.
FT    TRANSMEM   309    332
FT    DOMAIN     333    362            CYTOPLASMIC TAIL.
FT    DISULFID   125    188            BY SIMILARITY.
FT    DISULFID   227    283            BY SIMILARITY.
FT    CARBOHYD   110    110            BY SIMILARITY.
SQ    SEQUENCE   362 AA;  40204 MW;  3CF119AD CRC32;

Query Match             79.6%;  Score 39;  DB 1;  Length 362;
Best Local Similarity   80.0%;  Pred. No. 1.2;
Matches    8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 REDLRILLRY 10
          ||:||:|||
Db     99 RENLRIALRY 108


RESULT 10
1B03_GORGO
ID   1B03_GORGO   STANDARD;      PRT;    362 AA.
AC   P30381;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN, GOGO-B0103 ALPHA CHAIN PRECURSOR.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92078860.
RA   LAWLOR D.A., WARREN E., TAYLOR P., PARHAM P.;
RT   "Gorilla class I major histocompatibility complex alleles: comparison
RT   to human and chimpanzee class I.";
RL   J. Exp. Med. 174:1491-1509(1991).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------
DR   EMBL; X60254; CAA42806.1; -.
DR   PIR; JH0541; JH0541.
DR   HSSP; P03989; 1HSA.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1     24            BY SIMILARITY.
FT   CHAIN       25    362            CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                    GOGO-B0103 ALPHA CHAIN.
FT   DOMAIN      25    114            EXTRACELLULAR ALPHA-1.
FT   DOMAIN     115    206            EXTRACELLULAR ALPHA-2.
FT   DOMAIN     207    298            EXTRACELLULAR ALPHA-3.
FT   DOMAIN     299    308            CONNECTING PEPTIDE.
FT   TRANSMEM   309    332
FT   DOMAIN     333    362            CYTOPLASMIC TAIL.
FT   DISULFID   125    188            BY SIMILARITY.
FT   DISULFID   227    283            BY SIMILARITY.
FT   CARBOHYD   110    110            BY SIMILARITY.
SQ   SEQUENCE   362 AA;  40248 MW;  FEA6A941 CRC32;

Query Match             79.6%;  Score 39;  DB 1;  Length 362;
Best Local Similarity   80.0%;  Pred. No. 1.2;
Matches    8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 REDLRILLRY 10
          ||:||:|||
Db     99 RENLRIALRY 108


RESULT 11
1B15_HUMAN
ID   1B15_HUMAN   STANDARD;      PRT;    362 AA.
AC   P10317;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2702 ALPHA CHAIN
DE   PRECURSOR (B-27K) (B27.2).
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86220133.
RA   SEEMANN G.H.A., REIN R.S., BROWN C.S., PLOEGH H.L.;
RT   "Gene conversion-like mechanisms may generate polymorphism in human
RT   class I genes.";
RL   EMBO J. 5:547-552(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   PARHAM P., ARNETT K.L., ADAMS E.J.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 86-107 AND 171-181.
RX   MEDLINE; 86042671.
RA   VEGA M.A., EZQUERRA A., ROJO S., APARICIO P., BRAGADO R.,
RA   LOPEZ DE CASTRO J.A.;
RT   "Structural analysis of an HLA-B27 functional variant: identification
RT   of residues that contribute to the specificity of recognition by
RT   cytolytic T lymphocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7394-7398(1985).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------
DR   EMBL; X03664; CAA27301.1; -.
DR   EMBL; X03667; CAA27301.1; JOINED.
DR   EMBL; L38504; AAA69724.1; -.
DR   PIR; B25092; HLHUBK.
DR   HSSP; P03989; 1HSA.
```

```
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                B-27 B*2702 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    308       CONNECTING PEPTIDE.
FT   TRANSMEM    309    332
FT   DOMAIN      333    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE    362 AA;  40397 MW;  9798F0BB CRC32;

Query Match             79.6%;  Score 39;  DB 1;  Length 362;
Best Local Similarity   80.0%;  Pred No. 1.2;
Matches    8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        1 REDLRILLRY 10
            ||:|||:|||
Db       99 RENLRIALRY 108

RESULT 12
1B47_HUMAN
ID   1B47_HUMAN     STANDARD;      PRT;   362 AA.
AC   P30487;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-49(B-21) B*4901 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89235215.
RA   PARHAM P., LAWLOR D.A., LOMEN C.E., ENNIS P.D.;
RT   "Diversity and diversification of HLA-A,B,C alleles.";
RL   J. Immunol. 142:3937-3950(1989).
RN   [2]
RP   REVISION TO 78.
RX   MEDLINE; 93056529.
RA   HILDEBRAND W.H., MADRIGAL J.A., BELICH M.P., ZEMMOUR J., WARD F.E.,
RA   WILLIAMS R.C., PARHAM P.;
RT   "Serologic cross-reactivities poorly reflect allelic relationships in
RT   the HLA-B12 and HLA-B21 groups. Dominant epitopes of the alpha 2
RT   helix.";
RL   J. Immunol. 149:3563-3568(1992).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M24037; AAA02950.1; -.
DR   HSSP; P30491; 1A1M.
DR   MIM; 142830; -.
```

```
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                B-49(B-21) B*4901 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    309       CONNECTING PEPTIDE.
FT   TRANSMEM    310    333
FT   DOMAIN      334    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE    362 AA;  40581 MW;  E996F82F CRC32;

Query Match             79.6%;  Score 39;  DB 1;  Length 362;
Best Local Similarity   80.0%;  Pred No. 1.2;
Matches    8;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        1 REDLRILLRY 10
            ||:|||:|||
Db       99 RENLRIALRY 108

RESULT 13
1B49_HUMAN
ID   1B49_HUMAN     STANDARD;      PRT;   362 AA.
AC   P18464;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-51(B-5) B*5101 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90207291.
RA   ENNIS P.D., ZEMMOUR J., SALTER R.D., PARHAM P.;
RT   "Rapid cloning of HLA-A,B cDNA by using the polymerase chain
RT   reaction: frequency and nature of errors produced in amplification.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2833-2837(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89080265.
RA   HAYASHI H., ENNIS P.D., ARIGA H., SALTER R.D., PARHAM P., KANO K.,
RA   TAKIGUCHI M.;
RT   "HLA-B51 and HLA-Bw52 differ by only two amino acids which are in the
RT   helical region of the alpha 1 domain.";
RL   J. Immunol. 142:306-311(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89233295.
RA   POHLA H., KUON W., TABACZEWSKI P., DOERNER C., WEISS E.H.;
RT   "Allelic variation in HLA-B and HLA-C sequences and the evolution of
RT   the HLA-B alleles.";
RL   Immunogenetics 29:297-307(1989).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M32319; AAA36232.1; -.
DR   EMBL; M22792; AAA59620.1; ALT_SEQ.
DR   EMBL; M22786; AAA59620.1; JOINED.
DR   EMBL; M22787; AAA59620.1; JOINED.
DR   EMBL; M22788; AAA59620.1; JOINED.
DR   EMBL; M22789; AAA59620.1; JOINED.
DR   EMBL; M22790; AAA59620.1; JOINED.
DR   EMBL; M22791; AAA59620.1; JOINED.
DR   EMBL; L41087; AAA64513.1; -.
DR   EMBL; L41086; AAA64513.1; JOINED.
DR   PIR; A30345; A30345.
DR   PIR; A30548; A30548.
DR   HSSP; P30491; 1A1M.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL   1   24
FT   CHAIN   25   362   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT   B-51(B-5) B*5101 ALPHA CHAIN.
FT   DOMAIN   25   114   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   115   206   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   207   298   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   299   308   CONNECTING PEPTIDE.
FT   TRANSMEM   309   332
FT   DOMAIN   333   362   CYTOPLASMIC TAIL.
FT   CARBOHYD   110   110   BY SIMILARITY.
FT   DISULFID   125   188   BY SIMILARITY.
FT   DISULFID   227   283   BY SIMILARITY.
SQ   SEQUENCE   362 AA;   40566 MW;   4D846F30 CRC32;

Query Match   79.6%;   Score 39;   DB 1;   Length 362;
Best Local Similarity   80.0%;   Pred. No. 1.2;
Matches   8;   Conservative   1;   Mismatches   1;   Indels   1;   Gaps   0;

Qy   1   REDLRLLRY   10
    ||:||:||| |||
Db   99   RENLRIALRY   108

RESULT 14
1B52_HUMAN
ID   1B52_HUMAN   STANDARD;   PRT;   362 AA.
AC   P30489;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-51(B-5) B*5104 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92269955.
RA   BELICH M.P., MADRIGAL J.A., HILDEBRAND W.H., ZEMMOUR J.,
RA   WILLIAMS R.C., LUZ R., PETZL-ERLER M.L., PARHAM P.;
RT   "Unusual HLA-B alleles in two tribes of Brazilian Indians.";
RL   Nature 357:326-329(1992).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its

CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z15143; CAA78849.1; -.
DR   HSSP; P30491; 1A1M.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL   1   24
FT   CHAIN   25   362   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT   B-51(B-5) B*5104 ALPHA CHAIN.
FT   DOMAIN   25   114   EXTRACELLULAR ALPHA-1.
FT   DOMAIN   115   206   EXTRACELLULAR ALPHA-2.
FT   DOMAIN   207   298   EXTRACELLULAR ALPHA-3.
FT   DOMAIN   299   308   CONNECTING PEPTIDE.
FT   TRANSMEM   309   332
FT   DOMAIN   333   362   CYTOPLASMIC TAIL.
FT   CARBOHYD   110   110   BY SIMILARITY.
FT   DISULFID   125   188   BY SIMILARITY.
FT   DISULFID   227   283   BY SIMILARITY.
SQ   SEQUENCE   362 AA;   40560 MW;   F22F08AB CRC32;

Query Match   79.6%;   Score 39;   DB 1;   Length 362;
Best Local Similarity   80.0%;   Pred. No. 1.2;
Matches   8;   Conservative   1;   Mismatches   1;   Indels   1;   Gaps   0;

Qy   1   REDLRLLRY   10
    ||:||:||| |||
Db   99   RENLRIALRY   108

RESULT 15
1B53_HUMAN
ID   1B53_HUMAN   STANDARD;   PRT;   362 AA.
AC   P30490;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-52(B-5) B*5201 ALPHA CHAIN
DE   PRECURSOR.
GN   HLA-B OR HLAB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89080265.
RA   HAYASHI H., ENNIS P.D., ARIGA H., SALTER R.D., PARHAM P., KANO K.,
RA   TAKIGUCHI M.;
RT   "HLA-B51 and HLA-Bw52 differ by only two amino acids which are in the
RT   helical region of the alpha 1 domain.";
RL   J. Immunol. 142:306-311(1989).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC   THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC   MICROGLOBULIN).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M22799; AAA59645.1; ALT_SEQ.
DR   EMBL; M22793; AAA59645.1; JOINED.
DR   EMBL; M22794; AAA59645.1; JOINED.

```
DR   EMBL; M22795; AAA59645.1; JOINED.
DR   EMBL; M22796; AAA59645.1; JOINED.
DR   EMBL; M22797; AAA59645.1; JOINED.
DR   EMBL; M22798; AAA59645.1; JOINED.
DR   PIR; B30345; B30345.
DR   PIR; B30548; B30548.
DR   HSSP; P30491; 1A1M.
DR   MIM; 142830; -.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     24
FT   CHAIN        25    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                BW-52(B-5) B*5201 ALPHA CHAIN.
FT   DOMAIN       25    114       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      115    206       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      207    298       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      299    308       CONNECTING PEPTIDE.
FT   TRANSMEM    309    332
FT   DOMAIN      333    362       CYTOPLASMIC TAIL.
FT   CARBOHYD    110    110       BY SIMILARITY.
FT   DISULFID    125    188       BY SIMILARITY.
FT   DISULFID    227    283       BY SIMILARITY.
SQ   SEQUENCE    362 AA;  40521 MW;  3B436FE8 CRC32;

Query Match             79.6%;  Score 39;  DB 1;  Length 362;
Best Local Similarity   80.0%;  Pred. No. 1.2;
Matches    8;  Conservative   1;  Mismatches   1;  Indels   1;  Gaps   0;

Qy      1 REDLRILLRY 10
          !!:!!!|||
Db     99 RENLRIALRY 108

Search completed: February  8, 2000, 00:59:53
Job time: 3782 sec
```

THIS PAGE BLANK (USPTO)

```
          GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        February  8, 2000, 13:17:41 ; Search time 209.03 Seconds
                                             (without alignments)
                                             3.317 Million cell updates/sec

Title:          US-08-653-294-14
Perfect score:  49
Sequence:       1 REDLRILLRY 10

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters:    225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :  SPTREMBL_12:*
             1:  sp_archea:*
             2:  sp_bacteria:*
             3:  sp_fungi:*
             4:  sp_human:*
             5:  sp_invertebrate:*
             6:  sp_mammal:*
             7:  sp_mhc:*
             8:  sp_organelle:*
             9:  sp_phage:*
            10:  sp_plant:*
            11:  sp_rodent:*
            12:  sp_virus:*
            13:  sp_vertebrate:*
            14:  sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	39	7	O19688	O19688 homo sapien
2	44	89.8	90	7	O19193	O19193 homo sapien
3	44	89.8	181	7	O62898	O62898 homo sapien
4	44	89.8	181	7	O78138	O78138 homo sapien
5	44	89.8	181	7	O78142	O78142 homo sapien
6	44	89.8	274	7	O19692	O19692 homo sapien
7	44	89.8	322	7	O19627	O19627 homo sapien
8	44	89.8	359	7	O29934	O29934 homo sapien
9	44	89.8	362	7	O29705	O29705 homo sapien
10	44	89.8	362	7	O29846	O29846 homo sapien
11	44	89.8	362	7	O78189	O78189 homo sapien
12	44	89.8	89	7	O19569	O19569 homo sapien
13	39	79.6	90	7	O46697	O46697 gorilla gor
14	39	79.6	133	7	O19189	O19189 homo sapien
15	39	79.6	137	7	Q95533	Q95533 pan troglod
16	39	79.6	138	7	O78209	O78209 homo sapien
17	39	79.6	172	7	O19770	O19770 homo sapien
18	39	79.6	172	7	O19774	O19774 homo sapien
19	39	79.6	172	7	O19775	O19775 homo sapien
20	39	79.6	172	7	O19780	O19780 homo sapien
21	39	79.6	172	7	Q95364	Q95364 homo sapien
22	39	79.6	172	7	O19771	O19771 homo sapien
23	39	79.6	172	7	O19772	O19772 homo sapien
24	39	79.6	175	7	O19773	O19773 homo sapien
25	39	79.6	180	7	Q29694	Q29694 homo sapien
26	39	79.6	180	7	O19607	O19607 homo sapien
27	39	79.6	180	7	O19608	O19608 homo sapien
28	39	79.6	180	7	O19609	O19609 homo sapien
29	39	79.6	180	7	O19610	O19610 homo sapien
30	39	79.6	180	7	O19611	O19611 homo sapien
31	39	79.6	180	7	O19612	O19612 homo sapien
32	39	79.6	180	7	O19613	O19613 homo sapien
33	39	79.6	181	7	O46703	O46703 homo sapien
34	39	79.6	181	7	O62917	O62917 homo sapien
35	39	79.6	181	7	O62892	O62892 homo sapien
36	39	79.6	181	7	O62899	O62899 homo sapien
37	39	79.6	181	7	O62920	O62920 homo sapien
38	39	79.6	181	7	O62922	O62922 homo sapien
39	39	79.6	181	7	O62923	O62923 homo sapien
40	39	79.6	181	7	O19623	O19623 homo sapien
41	39	79.6	181	7	O19747	O19747 homo sapien
42	39	79.6	181	7	Q29667	Q29667 homo sapien
43	39	79.6	181	7	Q30198	Q30198 homo sapien
44	39	79.6	181	7	Q29708	Q29708 homo sapien
45	39	79.6	181	7	Q29724	Q29724 homo sapien

ALIGNMENTS

```
RESULT  1
O19688
ID   O19688       PRELIMINARY;       PRT;    39 AA.
AC   O19688;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HLA-B27 VARIANT EXON 2 (ALPHA1 DOMAIN) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BLASCZYK R., WEBER M., SALAMA A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X83727; CAA58698.1; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER          1      1
FT   NON_TER         39     39
SQ   SEQUENCE    39 AA;   4748 MW;  6F714D4C CRC32;

Query Match             89.8%;  Score 44;  DB 7;  Length 39;
Best Local Similarity   90.0%;  Pred. No. 0.07;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy          1 REDLRILLRY 10
              |||| ||||
Db         24 REDLRTLLRY 33

RESULT  2
O19193
ID   O19193       PRELIMINARY;       PRT;    90 AA.
AC   O19193;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HISTOCOMPATIBILITY ANTIGEN ALPHA 1 DOMAIN (FRAGMENT).
GN   HLA-B27.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 9237445.
RA   HIGGINS C.M., LUND T., SHIPLEY M.E., EBRINGER A.,
RA   SADOWSKA-WROBLEWSKA M., CRAIG R.K.;
RT   "Ankylosing spondylitis and HLA-B27: restriction fragment length
RT   polymorphism and sequencing of an HLA-B27 allele from a patient with
RT   ankylosing spondylitis.";
RL   Ann. Rheum. Dis. 51:855-862(1992).
DR   EMBL; S39758; CAB27364.1; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER     90     90
SQ   SEQUENCE    90 AA;  10571 MW;  F22CB4E CRC32;

Query Match             89.8%;  Score 44;  DB 7;  Length 90;
Best Local Similarity   90.0%;  Pred. No. 0.17;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps     0;

Qy      1 REDLRILLRY 10
           |||| ||||
Db     75 REDLRTLLRY 84

RESULT 3
O62898
ID   O62898      PRELIMINARY;      PRT;    181 AA.
AC   O62898;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOSMAN C.A., HURLEY C.K.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF054012; AAC32563.1; -.
DR   EMBL; AF054011; AAC32563.1; JOINED.
DR   HSSP; P10318; 1ROG.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER    181    181
SQ   SEQUENCE   181 AA;  21107 MW;  D8E533DD CRC32;

Query Match             89.8%;  Score 44;  DB 7;  Length 181;
Best Local Similarity   90.0%;  Pred. No. 0.34;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps     0;

Qy      1 REDLRILLRY 10
           |||| ||||
Db     74 REDLRTLLRY 83

RESULT 4
O78138
ID   O78138      PRELIMINARY;      PRT;    181 AA.
AC   O78138;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;


OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KOSMAN C.A., HURLEY C.K.;
RT   "Novel HLA Class I B Locus Alleles.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071770; AAC35939.1; -.
DR   EMBL; AF071769; AAC35939.1; JOINED.
DR   HSSP; P10318; 1ROG.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER    181    181
SQ   SEQUENCE   181 AA;  21103 MW;  8CF468CF CRC32;

Query Match             89.8%;  Score 44;  DB 7;  Length 181;
Best Local Similarity   90.0%;  Pred. No. 0.34;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps     0;

Qy      1 REDLRILLRY 10
           |||| ||||
Db     74 REDLRTLLRY 83

RESULT 5
O78142
ID   O78142      PRELIMINARY;      PRT;    181 AA.
AC   O78142;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   STEINER N.K., HURLEY C.K., KOESTER R.P.;
RT   "Novel HLA-B allele.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF072764; AAC25779.1; -.
DR   EMBL; AF072763; AAC25779.1; JOINED.
DR   HSSP; P10318; 1ROG.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER    181    181
SQ   SEQUENCE   181 AA;  21079 MW;  2949B0F CRC32;

Query Match             89.8%;  Score 44;  DB 7;  Length 181;
Best Local Similarity   90.0%;  Pred. No. 0.34;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps     0;

Qy      1 REDLRILLRY 10
           |||| ||||
Db     74 REDLRTLLRY 83

RESULT 6
O19692
ID   O19692      PRELIMINARY;      PRT;    274 AA.
AC   O19692;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I HLA-B27 M (FRAGMENT).
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 87009855.
RA   COPPIN H.L., MCDEVITT H.O.;
RT   "Absence of polymorphism between HLA-B27 genomic exon sequences
RT   isolated from normal donors and ankylosing spondylitis patients.";
RL   J. Immunol. 137:2168-2172(1986).
DR   EMBL; M14013; AAA59643.1; -.
DR   HSSP; P10318; 1ROG.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER   1   1
FT   NON_TER   274   274
SQ   SEQUENCE   274 AA; 31659 MW; 9A74A6BA CRC32;

Query Match   89.8%; Score 44; DB 7; Length 274;
Best Local Similarity   90.0%; Pred. No. 0.53;
Matches   9; Conservative   0; Mismatches   1; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
              ||||| ||||
Db   75 REDLRTLLRY 84

RESULT   7
ID   O19627
ID   O19627;   PRELIMINARY;   PRT;   322 AA.
AC   O19627;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HLA-B37 (FRAGMENT).
GN   B-3701.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HURLEY C.K., BEI M., RODRIGUEZ S., JOHNSON A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U11267; AAA19927.1; -.
DR   HSSP; P30685; 1A9E.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER   322   322
SQ   SEQUENCE   322 AA; 36626 MW; DF3B7744 CRC32;

Query Match   89.8%; Score 44; DB 7; Length 322;
Best Local Similarity   90.0%; Pred. No. 0.62;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
              ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   8
ID   Q29934
ID   Q29934;   PRELIMINARY;   PRT;   359 AA.
AC   Q29934;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HLA-B.
GN   HLA-B27 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86149317.
RA   SZOTS H., RIETHMULLER G., WEISS E., MEO T.;
RT   "Complete sequence of HLA-B27 cDNA identified through the
RT   characterization of structural markers unique to the HLA-A, -B, and -C
RT   allelic series.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1428-1432(1986).
DR   EMBL; M12678; AAA59614.1; -.
DR   HSSP; P10318; 1ROG.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER   1   1
SQ   SEQUENCE   359 AA; 40042 MW; 069F7E64 CRC32;

Query Match   89.8%; Score 44; DB 7; Length 359;
Best Local Similarity   90.0%; Pred. No. 0.69;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
              ||||| ||||
Db   96 REDLRTLLRY 105

RESULT   9
ID   Q29705
ID   Q29705;   PRELIMINARY;   PRT;   362 AA.
AC   Q29705;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN HLA-B PRECURSOR.
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   BALAS A., SANTOS S., VICARIO J.L.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U31971; AAA98506.1; -.
DR   HSSP; P10318; 1ROG.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   Signal; MHC.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   362   MHC CLASS I ANTIGEN HLA-B.
SQ   SEQUENCE   362 AA; 40479 MW; 09C9D20A CRC32;

Query Match   89.8%; Score 44; DB 7; Length 362;
Best Local Similarity   90.0%; Pred. No. 0.7;
Matches   9; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 REDLRILLRY 10
              ||||| ||||
Db   99 REDLRTLLRY 108

RESULT   10
ID   Q29846
ID   Q29846;   PRELIMINARY;   PRT;   362 AA.
AC   Q29846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HUMAN LYMPHOCYTE ANTIGEN HLA-B27.

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE: 94375872.
RA   DEL PORTO P., D'AMATO M., FIORILLO M.T., TUOSTO L., PICCOLELLA E.,
RA   SORRENTINO R.;
RT   "Identification of a novel HLA-B27 subtype by restriction analysis of
RT   a cytotoxic gamma delta T cell clone.";
RL   J. Immunol. 153:3093-3100(1994).
DR   EMBL; Z33453; CAA83876.1; -.
DR   HSSP; P10318; 1ROG.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   362 AA;  40450 MW;  CCA23A50 CRC32;

Query Match             89.8%;
Best Local Similarity   90.0%;  Pred. No. 0.7;
Matches    9;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy      1 REDLRILLRY 10
            ||||| ||||
Db     99 REDLRTLLRY 108

RESULT 11
078189
ID   078189;      PRELIMINARY;       PRT;   362 AA.
AC   078189;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I ANTIGEN.
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SEURYNCK K.L., BAXTER-LOWE L.A.;
RT   "B27052 W496D.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF026218; AAC42275.1; -.
DR   HSSP; P10318; 1ROG.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
SQ   SEQUENCE   362 AA;  40486 MW;  2B0EF602 CRC32;

Query Match             89.8%;
Best Local Similarity   90.0%;  Pred. No. 0.7;
Matches    9;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy      1 REDLRILLRY 10
            ||||| ||||
Db     99 REDLRTLLRY 108

RESULT 12
019569
ID   019569;      PRELIMINARY;       PRT;   89 AA.
AC   019569;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).

GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CAO K., BURDETT L., ZHANG G., FERNANDEZ-VINA M.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017320; AAB70286.2; -.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER     89     89
SQ   SEQUENCE   89 AA;  10606 MW;  99D11089 CRC32;

Query Match             79.6%;
Best Local Similarity   80.0%;  Pred. No. 1.6;
Matches    8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 REDLRILLRY 10
            ||:||| |||
Db     74 RENLRIALRY 83

RESULT 13
046697
ID   046697;      PRELIMINARY;       PRT;   90 AA.
AC   046697;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MHC CLASS I ANTIGEN HLA-H ORTHOLOG (FRAGMENT).
GN   HLA-H.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHAMBA;
RA   GRIMSLEY C., MATHER K.A., OBER C.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022172; AAC99794.1; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER      1      1
FT   NON_TER     90     90
SQ   SEQUENCE   90 AA;  10689 MW;  5E5F2495 CRC32;

Query Match             79.6%;
Best Local Similarity   80.0%;  Pred. No. 1.6;
Matches    8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 REDLRILLRY 10
            ||:||| |||
Db     75 RENLRIALRY 84

RESULT 14
019189
ID   019189;      PRELIMINARY;       PRT;   133 AA.
AC   019189;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MHC CLASS I HISTOCOMPATIBILITY ANTIGEN-B (HLA-B-27KSH) (FRAGMENT).
GN   HLA-B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
```

```
RA   PETERSDORF E.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U18659; AAB60357.1; -.
DR   MIM; 142830; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC I.
FT   NON_TER        1        1
FT   NON_TER      133      133
SQ   SEQUENCE     133 AA;    15491 MW;   3A3BC802 CRC32;

Query Match           79.6%;   Score 39;  DB 7;  Length 133;
Best Local Similarity   80.0%;   Pred. No. 2.4;
Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     1 REDLRILLRY 10
         ||:||| |||
Db    27 RENLRIALRY 36


RESULT  15
Q95533
ID   Q95533         PRELIMINARY;          PRT;     137 AA.
AC   Q95533;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CLASS I HISTOCOMPATIBILITY ANTIGEN (FRAGMENT).
GN   HLA-B.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WODKA;
RX   MEDLINE; 94286544.
RA   MCADAM S.N., BOYSON J.E., LIU X., GARBER T.L., HUGHES A.L.,
RA   BONTROP R.E., WATKINS D.I.;
RT   "A uniquely high level of recombination at the HLA-B locus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:5893-5897(1994).
DR   EMBL; U05585; AAA50188.1; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
FT   NON_TER        1        1
FT   NON_TER      137      137
SQ   SEQUENCE     137 AA;    15922 MW;   B316D3BC CRC32;

Query Match           79.6%;   Score 39;  DB 7;  Length 137;
Best Local Similarity   80.0%;   Pred. No. 2.4;
Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     1 REDLRILLRY 10
         ||:|| ||||
Db    40 RENLRTLLRY 49


Search completed: February  8, 2000, 13:17:41
Job time: 32490 sec
```

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-14 to: GenEmbl:*   out_format : pfs
Date: Feb 8, 2000 4:40 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgnl_1/USPTO_spool/US08653294/runat_04022000_160701_15779/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information.block:
Query: US-08-653-294-14
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 11370.480000

score_list:

sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pr1:HSLAB27V2	+	44.00	170.85	0.1709	195	! X83727 H.sapiens HLA-B27 variar
gb_pr3:HSAB27HLA1	+	44.00	168.07	0.2439	270	! AF072763 Homo sapiens MHC class
gb_pr3:HSB1524V1	+	44.00	168.07	0.2439	270	! AF054011 Homo sapiens MHC class
gb_pr3:HSMHCLBC1	+	44.00	168.07	0.2439	270	! AF071769 Homo sapiens MHC class C
gb_pr4:HS27052B1	+	44.00	168.07	0.2439	270	! AF102563 Homo sapiens MHC class
gb_pr4:HUMHLAWMBY2	+	44.00	168.07	0.2439	270	! AF110257 Homo sapiens MHC class
gb_pr2:HUMB27052G	+	44.00	162.10	0.5268	546	! L76095 Homo sapiens MHC class I
gb_pr1:HUMMHBM2	+	44.00	158.58	0.8239	822	! M14013 Human MHC class I HLA-B2
gb_pr1:HSU11267	+	44.00	157.19	0.9852	968	! U11267 Human HLA-B37 (B-3701) m
gb_pat:A28264	+	44.00	156.69	1.04	1017	! M62852 Human MHC class I HLA-B
gb_pat:E01342	+	44.00	156.69	1.05	1026	! A28264 H.sapiens mRNA for HLA-
gb_pr1:HUMMHZUNIA	+	44.00	156.22	1.11	1026	! E01342 cDNA encoding C-termina
gb_pr2:HSU31971	+	44.00	156.19	1.11	1084	! L20086 Human MHC class I (HLA-
gb_pr3:AF026218	+	44.00	156.19	1.12	1089	! U31971 Human MHC class I antig
gb_pr1:HUMHLABW4	+	44.00	156.15	1.12	1089	! AF026218 Homo sapiens MHC clas
gb_pr1:HUMMHCAG	+	44.00	156.15	1.13	1093	! M19756 Human MHC class I HLA-B
gb_pr1:HSHLAW1	+	44.00	155.07	1.13	1093	! M32320 Human MHC HLA protein,
gb_pr1:HSHLAB27X	+	44.00	154.88	1.29	1241	! X03665 Human class I MHC gene
gb_pr1:HUMMHB27A	+	44.00	153.42	1.33	1272	! Z33453 H.sapiens mRNA for huma
gb_pr1:HUMMHB27D	+	44.00	145.88	1.60	1507	! M12678 Human HLA-B27 mRNA, com
gb_pat:E01341	+	44.00	145.06	4.20	3649	! M54883 Human MHC class I HLA-B
gb_pr1:HUMMHB27B	+	44.00	145.06	4.67	4015	! E01341 Genomic DNA encoding HL
gb_pr4:S39758	+	44.00	144.59	4.67	4016	! M12967 Human MHC class I HLA-B
gb_pr1:HSHLAB27	-	44.00	144.50	4.96	4242	! S39758 HLA-B27 (HLA-B*2705)-hi
gb_pat:AR008238	+	44.00	140.89	5.01	4287	! X03945 Human gene for HLA-B27
gb_ba1:D90899	+	42.00	106.88	7.97	6553	! AR008238 Sequence 1 from paten
gb_ba1:PRM7GNHOM	+	40.00	124.57	624.87	133859	! D90899 Synechocystis sp. PCC
gb_htg6:AC016616	+	40.00	114.59	232.40	20485	! M58352 Proteus mirabilis 60k-r
gb_htg5:AC013786	+	40.00	104.29	871.32	68611	! AC016616 Homo sapiens chromoso
gb_pl1:ATF10M6	-	40.00	101.85	1.2e+03	91344	! AC013786 Homo sapiens chromoso
gb_pr3:HSBA46E17	-	40.00	101.66	1.2e+03	93353	! AL021811 Arabidopsis thaliana
gb_htg3:AC009409	-	40.00	100.29	1.5e+03	109710	! AL050402 Human DNA sequence
gb_htg7:AC013285	-	40.00	96.27	2.4e+03	175653	! AC009409 Homo sapiens clone
gb_pr1:HSHLABHBA	+	40.00	95.24	2.8e+03	198303	! AC013285 Homo sapiens clone
gb_pr1:HSHLABHBB	+	39.00	148.01	3.20	250	! Y08692 H.sapiens HLA-B gene, ex
gb_pr1:HSHLABHBD	+	39.00	148.01	3.20	250	! Y08693 H.sapiens HLA-B gene, ex
gb_pr2:HSHLABB1	+	39.00	148.01	3.20	250	! Y08694 H.sapiens HLA-B gene, ex
gb_pr2:HSHLABD1	+	39.00	148.01	3.20	250	! U90420 Human cell line THAI DCH
gb_pr2:HSHLABF1	+	39.00	148.01	3.20	250	! U90613 Human cell line THAI DCH
gb_pr2:HSHLABG1	+	39.00	148.01	3.20	250	! U90615 Human cell line THAI DCH
gb_pr2:HSHLABI1	+	39.00	148.01	3.20	250	! U90422 Human cell line THAI
gb_pr2:HSHLABJ1	+	39.00	148.01	3.20	250	! U90424 Human cell line THAI
gb_pr2:HSHLABT1	+	39.00	148.01	3.20	250	! U90418 Human cell line THAI

seq_name: gb_pr1:HSLAB27V2

seq_documentation_block:
LOCUS   HSLAB27V2   195 bp   DNA   PRI   31-JAN-1995
DEFINITION H.sapiens HLA-B27 variant gene (exon 2).
ACCESSION   X83727
VERSION   X83727.1 GI:663002
KEYWORDS   HLA-B gene; human leukocyte antigen; major histocompatability
            complex class I.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 195)
AUTHORS   Blasczyk,R., Weber,M. and Salama,A.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 195)
AUTHORS   Blasczyk,R.
TITLE   Direct Submission
JOURNAL   Submitted (06-JAN-1995) R. Blasczyk, Bloodbank,Dept.of Intern.
            Medicine, Div of Hematol. and Oncolog., Spandauer Damm 130, Univ.
            Hosp. Rudolf Virchow, Freie Univ., D- 14050 Berlin, FRG
FEATURES   Location/Qualifiers
source   1..195
                /organism="Homo sapiens"
                /isolate="1235"
                /db_xref="taxon:9606"
                /chromosome="6"
                /cell_type="EBV transformed cell line"
                /map="6p21.3"
CDS   <76..>195
                /codon_start=3
                /product="HLA-B27 variant exon 2 (alpha1 domain)"
                /protein_id="CAA58698.1"
                /db_xref="GI:663003"
                /db_xref="SPTREMBL:O19688"
                /translation="IEEGPEYWDRETQICKAKAQTDREDLRTLLRYNQSEA"
BASE COUNT   45 a   55 c   73 g   22 t
ORIGIN

alignment_scores:
         Quality: 44.00   Length: 10
         Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x HSLAB27V2

Align seg 1/1 to: HSLAB27V2   from: 1 to: 195

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||||| |||||||||||||||
  147 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 176

seq_name: gb_pr3:HSAB27HLA1

seq_documentation_block:
LOCUS   HSAB27HLA1   270 bp   DNA   PRI   08-JUL-1998
DEFINITION Homo sapiens MHC class I antigen HLA-B gene (HLA-B*27 variant
            allele), exon 2.
ACCESSION   AF072763
VERSION   AF072763.1 GI:3293562
KEYWORDS   .
SEGMENT   1 of 2
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

Percent Similarity: 90.000   Percent Identity: 90.000

```
alignment_block:
US-08-653-294-14 x HSB1524V1

Align seg 1/1  to: HSB1524V1  from: 1  to: 270

         1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
           |||||||||||||||  ||||||||||||
       222 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 251

seq_name: gb_pr3:HSMHCHLBC1

seq_documentation_block:
LOCUS       HSMHCHLBC1         270 bp    DNA          PRI    17-SEP-1998
DEFINITION  Homo sapiens MHC class I antigen HLA-B gene (HLA-B*5303 allele),
            exon 2.
ACCESSION   AF071769
VERSION     AF071769.1  GI:3243269
KEYWORDS
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Novel HLA Class I B Locus Alleles
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUN-1998) Microbiology & Immunology, Georgetown
            University, 3970 Reservoir Rd. N.W., Washington, DC 20007, USA
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="Homo sapiens"
                     /isolate="GN00231"
                     /db_xref="taxon:9606"
     exon            1..270
                     /gene="HLA-B"
                     /number=2
BASE COUNT       59 a       84 g       90 c       37 t
ORIGIN

alignment_scores:
        Quality:    44.00                Length:      10
        Ratio:      4.889                Gaps:          0
Percent Similarity: 90.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x HSMHCHLBC1

Align seg 1/1  to: HSMHCHLBC1  from: 1  to: 270

         1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
           |||||||||||||||  ||||||||||||
       222 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 251

seq_name: gb_pr4:HS27052B1

seq_documentation_block:
LOCUS       HS27052B1          270 bp    DNA          PRI    21-JUN-1999
DEFINITION  Homo sapiens MHC class I antigen HLA-B gene, HLA-B*2716 allele,
            exon 2.
ACCESSION   AF102563
VERSION     AF102563.1  GI:4704574
KEYWORDS
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
REFERENCE   1  (bases 1 to 270)
AUTHORS     Steiner,N.K., Hurley,C.K. and Koester,R.P.
TITLE       Novel-HLA-B allele
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 270)
AUTHORS     Steiner,N.K., Hurley,C.K. and Koester,R.P.
TITLE       Direct Submission
JOURNAL     Submitted (21-JUN-1998) Microbiology and Immunology, Georgetown
            University Medical Center, 3970 Reservoir Road NW, Washington, DC
            20007, USA
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="Homo sapiens"
                     /isolate="GN00211"
                     /db_xref="taxon:9606"
     exon            1..270
                     /gene="HLA-B"
                     /number=2
BASE COUNT       54 a       93 g       87 c       36 t
ORIGIN

alignment_scores:
        Quality:    44.00                Length:      10
        Ratio:      4.889                Gaps:          0
Percent Similarity: 90.000    Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x HSAB27HLA1

Align seg 1/1  to: HSAB27HLA1  from: 1  to: 270

         1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
           ||||||||||||||| ||||||||||||
       222 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 251

seq_name: gb_pr3:HSB1524V1

seq_documentation_block:
LOCUS       HSB1524V1          270 bp    DNA          PRI    22-AUG-1998
DEFINITION  Homo sapiens isolate GN00211 MHC class I antigen HLA-B gene (B*1543
            allele), exon 2.
ACCESSION   AF054011
VERSION     AF054011.1  GI:2984766
KEYWORDS
SEGMENT     1 of 2
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Novel HLA Class I B locus alleles
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAR-1998) Microbiology & Immunology, Georgetown
            University, 3970 Reservoir Rd. NW, Washington, DC 20007, USA
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="Homo sapiens"
                     /isolate="GN00211"
                     /db_xref="taxon:9606"
     exon            1..270
                     /gene="HLA-B"
                     /number=2
BASE COUNT       57 a       86 g       90 c       37 t
ORIGIN

alignment_scores:
        Quality:    44.00                Length:      10
        Ratio:      4.889                Gaps:          0
```

```
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Novel Class I HLA-B Alleles
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 270)
AUTHORS     Kosman,C.A. and Hurley,C.K.
TITLE       Direct Submission
JOURNAL     Submitted (28-OCT-1998) Microbiology and Immunology, Georgetown
            University, 3970 Reservoir Rd. N.W., Washington, DC 20007, USA
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="Homo sapiens"
                     /isolate="GN00246"
                     /db_xref="taxon:9606"
     exon            1..270
                     /gene="HLA-B"
                     /number=2
BASE COUNT     56 a     88 c     90 g     36 t
ORIGIN


seq_name: gb_pr4:HSHLAWMBY2

alignment_scores:
        Quality:    44.00                  Length:    10
        Ratio:       4.889                 Gaps:       0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HS27052B1

Align seg 1/1 to: HS27052B1 from:  1 to: 270

      1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
        |||||||||||  ||||||||||||
    222 CGAGAGGACCTGCGGGACCCTGCTCCGCTAC 251


seq_documentation_block:
LOCUS       HSHLAWMBY2    270 bp    DNA          PRI    06-APR-1999
DEFINITION  Homo sapiens MHC class I antigen HLA-B gene (HLA-B*27 allele), exon
            2.
ACCESSION   AF110257
VERSION     AF110257.1  GI:4566542
KEYWORDS    .
SEGMENT     2 of 3
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270)
AUTHORS     Wu,J., Bassinger,S., Montoya,G.D., Yee,J., Griffith,B.B.,
            Kearns,J., McKeen,M., Birkos,S., Kamoun,M. and Williams,T.M.
TITLE       Identification of new HLA-B alleles in potential bone marrow donors
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 270)
AUTHORS     Wu,J., Bassinger,S., Montoya,G.D., Yee,J., Griffith,B.B.,
            Kearns,J., McKeen,M., Birkos,S., Kamoun,M. and Williams,T.M.
TITLE       Direct Submission
JOURNAL     Submitted (30-NOV-1998) Pathology, Univ. New Mexico, 915 Camino de
            Salud, NE, Albuquerque, NM 87131, USA
FEATURES             Location/Qualifiers
     source          1..270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p23"
                     /chromosome="6"
     exon            1..270
                     /gene="HLA-B"
                     /number=2
BASE COUNT     54 a     87 c     93 g     36 t
ORIGIN
```

```
alignment_scores:
        Quality:    44.00                  Length:    10
        Ratio:       4.889                 Gaps:       0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HSHLAWMBY2

Align seg 1/1 to: HSHLAWMBY2 from:  1 to: 270

      1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
        |||||||||||  ||||||||||||
    222 CGAGAGGACCTGCGGGACCCTGCTCCGCTAC 251


seq_name: gb_pr2:HUMB27052G

seq_documentation_block:
LOCUS       HUMB27052G    546 bp    DNA          PRI    27-FEB-1996
DEFINITION  Homo sapiens MHC class I HLA-B*27052 gene, exons.
ACCESSION   L76095
VERSION     L76095.1  GI:1203957
KEYWORDS    cell surface antigen; cell surface glycoprotein; class I gene;
            integral membrane protein; major histocompatibility complex.
SOURCE      Homo sapiens DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 546)
AUTHORS     Marcos,C.Y., Fernandez-Vina,M.A., Lazaro,A.M. and Stastny,P.
TITLE       Novel HLA-B Alleles
JOURNAL     Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..546
                     /gene="B-2705v"
     exon            1..270
                     /gene="B-2705v"
     exon            271..546
                     /gene="B-2705v"
     allele          454
                     /gene="B*27052"
                     /replace="t"
     gene            454
                     /gene="B*27052"
BASE COUNT    109 a    172 c    195 g     70 t
ORIGIN


alignment_scores:
        Quality:    44.00                  Length:    10
        Ratio:       4.889                 Gaps:       0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HUMB27052G

Align seg 1/1 to: HUMB27052G from:  1 to: 546

      1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
        |||||||||||  ||||||||||||
    222 CGAGAGGACCTGCGGGACCCTGCTCCGCTAC 251


seq_name: gb_pr1:HUMMHBM2

seq_documentation_block:
LOCUS       HUMMHBM2    822 bp    DNA          PRI    07-JAN-1995
DEFINITION  Human MHC class I HLA-B27 M+ gene, exons 2-4 (introns unsequenced).
ACCESSION   M14013
VERSION     M14013.1  GI:187743
KEYWORDS
```

```
SOURCE      Human DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 822)
AUTHORS     Coppin,H.L. and McDevitt,H.O.
TITLE       Absence of polymorphism between HLA-B27 genomic exon sequences
            isolated from normal donors and ankylosing spondylitis patients
JOURNAL     J. Immunol. 137 (7), 2168-2172 (1986)
MEDLINE     87009855
FEATURES             Location/Qualifiers
     source          1..822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
     gene            1..822
                     /gene="HLA-B"
     CDS             <1..>822
                     /gene="HLA-B"
                     /note="HLA-B27 M2+"
                     /codon_start=1
                     /db_xref="GDB:G00-120-048"
                     /protein_id="AAA59643.1"
                     /db_xref="GI:187744"
                     /translation="GSHSMRYFHTSVSRPGRGEPRFITVGYVDDTLFVRFDSDAASPR
                     EEPRAPWIEQEGPEYWDRETQICKAKAQTDREDLRTLLRYYNQSEAGSHTLQNMYGCD
                     VGPDGRLLRGYHQDAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYL
                     EGECVEWLRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQ
                     RDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRRW"
BASE COUNT      176 a    256 c    277 g    113 t
ORIGIN

alignment_scores:
        Quality:     44.00            Length:     10
        Ratio:        4.889            Gaps:        0
Percent Similarity:  90.000     Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HUMMHBM2

Align seg 1/1  to: HUMMHBM2  from: 1  to: 822

         1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
           ||||||||||||| |||||||||||
       223 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 252

seq_name: gb_pri:HSU11267

seq_documentation_block:
LOCUS       HSU11267       968 bp    mRNA            PRI       21-JUL-1994
DEFINITION  Human HLA-B37 (B-3701) mRNA, partial cds.
ACCESSION   U11267
VERSION     U11267.1  GI:511785
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Hurley,C.K., Bei,M., Rodriguez,S. and Johnson,A.
TITLE       HLA-B1
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 968)
AUTHORS     Hurley,C.K.
TITLE       Direct Submission
JOURNAL     Submitted (23-JUN-1994) Carolyn K. Hurley, Microbiology, Georgetown
            University School of Medicine, 3900 Reservoir Road, N.W.,
            Washington, D.C. 20007 USA
FEATURES             Location/Qualifiers
     source          1..968
                     /organism="Homo sapiens"
                     /isolate="GU2760"
```

```
                     /db_xref="taxon:9606"
                     /haplotype="HLA-A2, CW6, B37 / A-, CW7, B71"
                     /cell_line="B lymphoblastoid"
                     /chromosome="6"
     gene            1..968
                     /gene="B-3701"
     CDS             1..>968
                     /gene="B-3701"
                     /citation=[1]
                     /codon_start=1
                     /function="peptide presentation; histocompatibility
                     molecule"
                     /product="HLA-B37"
                     /protein_id="AAA19927.1"
                     /db_xref="GI:511786"
                     /translation="MRVTAPRTLLLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEP
                     RFISVGYVDDTQFVRFDSDAASPRTEPRAPWIEQEGPEYWDRETQISKTNTQTYREDL
                     RTLLRYYNQSEAGSHTIQRMSGCDVGPDGRLLRGYNQFAYDGKDYIALNEDLSSWTAA
                     DTAAQITQRKWEAARVAEQDRAYLEGTCVEWLRRYLENGKETLQRADPPKTHVTHHPI
                     SDHEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSG
                     EEQRYTCHVQHEGLPKPLTLRWEPSSQSTIPIVGIVAGLAVLAVVV"
BASE COUNT      196 a    310 c    314 g    148 t
ORIGIN

alignment_scores:
        Quality:     44.00            Length:     10
        Ratio:        4.889            Gaps:        0
Percent Similarity:  90.000     Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HSU11267

Align seg 1/1  to: HSU11267  from: 1  to: 968

         1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
           ||||||||||||| |||||||||||
       295 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 324

seq_name: gb_pri:HUMMHC

seq_documentation_block:
LOCUS       HUMMHC        1017 bp    mRNA            PRI       07-JAN-1995
DEFINITION  Human MHC class I HLA-B27-HS mRNA, 3' end.
ACCESSION   M62852
VERSION     M62852.1  GI:187760
KEYWORDS    cell surface antigen; class I gene; integral membrane protein;
            major histocompatibility complex.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1017)
AUTHORS     Choo,S.Y., Fan,L.A. and Hansen,J.A.
TITLE       A novel HLA-B27 allele maps B27 allospecificity to the region
            around position 70 in the alpha 1 domain
JOURNAL     J. Immunol. 147 (1), 174-180 (1991)
MEDLINE     91268545
FEATURES             Location/Qualifiers
     source          1..1017
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
     gene            1..1017
                     /gene="HLA-B"
     CDS             <1..1017
                     /gene="HLA-B"
                     /codon_start=1
                     /db_xref="GDB:G00-120-048"
                     /product="MHC HLA-B27-HS"
                     /protein_id="AAA59647.1"
                     /db_xref="GI:187761"
                     /translation="GSHSMRYFHTSVSRPGRGEPRFITVGYVDDTLFVRFDSDAASPR
```

```
                EEPRAPWIEQEGPEYWDRETQICKAKAQTDREDLRTLLRYNQSEAGSHTLQSMYGCD
                VGPDGRLLRGHNQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAARVAEQLRAYL
                EGECVEWLRRYLENGKETLQRADPPKTHVTHPISDHEATLRCWALGFYPAEITLTWQ
                RDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEP
                SSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDV
                SLTA"
BASE COUNT     207 a    308 c    343 g    159 t
ORIGIN

alignment_scores:
       Quality:       44.00               Length:       10
       Ratio:          4.889               Gaps:          0
Percent Similarity: 90.000     Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x HUMMHC    ..

Align seg 1/1   to: HUMMHC    from: 1  to: 1017

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||| |||| |||||||||||
  223 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 252

seq_name: gb_pat:A28264

seq_documentation_block:
LOCUS       A28264       1026 bp   DNA            PAT    24-MAY-1995
DEFINITION  H.sapiens mRNA for HLA-B 27 from patent EP0225069.
ACCESSION   A28264
VERSION     A28264.1  GI:905320
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1026)
AUTHORS     Szoets,H., Weiss,E., Doerner,C., Lang,M., Meo,T. and
            Riethmueller,G.
TITLE       HLA-B 27, DNA coding therefor and its utilization
JOURNAL     Patent: EP 0226069-A 1 24-JUN-1987;
            Riethmueller, Gert, Prof. Dr
FEATURES    Location/Qualifiers
     source         1..1026
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
BASE COUNT     213 a    307 c    344 g    162 t
ORIGIN

alignment_scores:
       Quality:       44.00               Length:       10
       Ratio:          4.889               Gaps:          0
Percent Similarity: 90.000     Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x A28264    ..

Align seg 1/1   to: A28264    from: 1  to: 1026

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||| |||| |||||||||||
  223 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 252

seq_name: gb_pat:E01342

seq_documentation_block:
LOCUS       E01342       1026 bp   RNA            PAT    29-SEP-1997
DEFINITION  cDNA encoding C-terminal Fragment of HLA-B27.
ACCESSION   E01342
VERSION     E01342.1  GI:2169599
KEYWORDS    JP 1987228281-A/2.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1026)
AUTHORS     Hanneroore,S., Erizabeeto,W., Kurisuta,D., Maagotsuto,R., Tomaso,M.
            and Geruto,R..
TITLE       HLA-B27, DNA ENCODING THE SAME AND ITS USE
JOURNAL     Patent: JP 1987228281-A 2 07-OCT-1987;
            BEHRINGWERKE AG
COMMENT     OS   Human
            PN   JP 1987228281-A/2
            PD   07-OCT-1987
            PF   28-NOV-1986 JP 1986284078
            PR   28-NOV-1985 DE 85    3542024, 21-DEC-1985 DE 85    3545576 PI
            HANNEROORE SUTSUBEETSU, ERIZABEETO WAISU, KURISUTA DERUNAA, PI
            MAAGOTSUTO RANGU, TOMASO MEO, GERUTO RIITOMIYURAA PC
            C12N15/00,C07H21/04,C12P21/00,C12Q1/68,G01N33/577//A61K39/00, PC
            C07K13/00,
            PC   C07K15/06,(C12P21/00,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   fragment_type: C-Terminal Fragment;
            CC   *source: cell_type=Leukocyte;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1..1026
            FT                   /product='C-terminal fragment(exon 2 to 7)of
            FT                   HLA-B27'.
FEATURES    Location/Qualifiers
     source          1..1026
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     213 a    307 c    344 g    162 t
ORIGIN

alignment_scores:
       Quality:       44.00               Length:       10
       Ratio:          4.889               Gaps:          0
Percent Similarity: 90.000     Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x E01342    ..

Align seg 1/1   to: E01342    from: 1  to: 1026

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||| |||| |||||||||||
  223 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 252

seq_name: gb_prl:HUMMHZUNIA

seq_documentation_block:
LOCUS       HUMMHZUNIA   1084 bp   mRNA           PRI    27-SEP-1993
DEFINITION  Human MHC class I (HLA-B 27052) mRNA fragment.
ACCESSION   L20086
VERSION     L20086.1  GI:307282
KEYWORDS    class I gene; lymphocyte antigen; major histocompatibility complex.
SOURCE      Homo sapiens (strain South American Amerindian) cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1084)
AUTHORS     Watkins,D.I., McAdam,S.N., Liu,X., Strang,C.R., Milford,E.L.,
            Levine,C.G., Garber,T.L., Dogon,A.L., Lord,C.I., Ghim,S.H.,
            Troup,G.M., Hughes,A.L. and Letvin,N.L.
TITLE       New recombinant HLA-B alleles in a tribe of South American
            Amerindians indicate rapid evolution of MHC class I loci
JOURNAL     Nature 357, 329-333 (1992)
MEDLINE     92269956
FEATURES    Location/Qualifiers
```

```
source          1..1084
                /organism="Homo sapiens"
                /strain="South American Amerindian"
                /db_xref="taxon:9606"
BASE COUNT       218 a    329 c    362 g    175 t
ORIGIN

alignment_scores:
        Quality:     44.00              Length:    10
          Ratio:    4.889                Gaps:     0
Percent Similarity:  90.000   Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HUMMHZUNIA    ..

Align seg 1/1 to: HUMMHZUNIA    from: 1 to: 1084

      1 ArgGluAspLeuArgIleLeuArgTyr 10
        |||||||||||||  |||||||||||||
    290 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 319

seq_name: gb_pr2:HSU31971

seq_documentation_block:
LOCUS       HSU31971    1089 bp    mRNA              PRI       27-APR-1996
DEFINITION  Human MHC class I antigen HLA-B precursor (HLA-B) mRNA, complete
            cds.
ACCESSION   U31971
VERSION     U31971.1    GI:1144333
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1089)
AUTHORS     Balas,A., Santos,S. and Vicario,J.L.
TITLE       Serological and molecular characterization of a novel HLA-B allele
JOURNAL     Unpublished (1995)
REFERENCE   2 (bases 1 to 1089)
AUTHORS     Balas,A.
TITLE       Direct Submission
JOURNAL     Submitted (20-JUL-1995) Antonio Balas, Centro de Transfusion de
            Madrid, Histocompatibility, Menendez Pelayo, 65, Madrid, 28009,
            Spain
FEATURES             Location/Qualifiers
     source          1..1089
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /haplotype="A1 Cw2 B- Bw4 DR11 DR52 DQ7"
                     /cell_line="CTM 8958127"
                     /cell_type="lymphocyte"
                     /tissue_type="blood"
                     /dev_stage="adult"
     sig_peptide     1..72
                     /gene="HLA-B"
     gene            1..1089
                     /gene="HLA-B"
                     /allele="HLA-B*3702"
     CDS             1..1089
                     /gene="HLA-B"
                     /codon_start=1
                     /product="MHC class I antigen HLA-B precursor"
                     /protein_id="AAA98506.1"
                     /db_xref="GI:1144334"
                     /translation="MRVTAPRTLLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEP
                     RFISVGYVDDTQFVRFDSDAASPRTEPRAPWIEQEGPEYWDRETQISKTNTQTYREDL
                     RTLLRYYNQSEAGSHTLQNMYGCDVGPDGRLLRGYHQDAYDGKDYIALNEDLSSWTAA
                     DTAAQITQRKWEAARVAEQLRAYLEGECVEWLRRYLENGKETIQRADPPKTHVTHHPI
                     SDHEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSG
                     EEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRR
                     KSSGGKGGSYSQAACSDSAQGSDVSLTA"

     exon            1..73
                     /gene="HLA-B"
                     /number=1
     mat_peptide     73..1086
                     /gene="HLA-B"
                     /product="MHC class I antigen HLA-B"
     misc_feature    73..342
                     /gene="HLA-B"
                     /note="encodes alpha-1 domain"
     exon            74..343
                     /gene="HLA-B"
                     /number=2
     misc_feature    343..618
                     /gene="HLA-B"
                     /note="encodes alpha-2 domain"
     exon            344..619
                     /gene="HLA-B"
                     /number=3
     misc_feature    619..894
                     /gene="HLA-B"
                     /note="encodes alpha-3 domain"
     exon            620..895
                     /gene="HLA-B"
                     /number=4
     misc_feature    895..1011
                     /gene="HLA-B"
                     /note="encodes transmembrane domain"
     exon            896..1012
                     /gene="HLA-B"
                     /number=5
     misc_feature    1012..1086
                     /gene="HLA-B"
                     /note="encodes cytoplasmic domain"
     exon            1013..1041
                     /gene="HLA-B"
                     /number=6
     exon            1042..1089
                     /gene="HLA-B"
                     /number=7
BASE COUNT       218 a    337 c    361 g    173 t
ORIGIN

alignment_scores:
        Quality:     44.00              Length:    10
          Ratio:    4.889                Gaps:     0
Percent Similarity:  90.000   Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x HSU31971    ..

Align seg 1/1 to: HSU31971    from: 1 to: 1089

      1 ArgGluAspLeuArgIleLeuArgTyr 10
        |||||||||||||  |||||||||||||
    295 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 324

seq_name: gb_pr3:AF026218

seq_documentation_block:
LOCUS       AF026218    1089 bp    mRNA              PRI       23-SEP-1998
DEFINITION  Homo sapiens MHC class I antigen HLA-B (HLA-B27052 allele) mRNA,
            complete cds.
ACCESSION   AF026218
VERSION     AF026218.1    GI:3543696
KEYWORDS    
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1089)
AUTHORS     Seurynck,K.L. and Baxter-Lowe,L.A.
TITLE       B27052 W496D
```

```
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1089)
AUTHORS      Seurynck,K.L. and Baxter-Lowe,L.A.
TITLE        Direct Submission
JOURNAL      Submitted (22-SEP-1997) Molecular Genetics, Richland Memorial
             Hospital, 7 Richland Medical Park, Columbia, SC 29203, USA
FEATURES             Location/Qualifiers
     source          1..1089
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /cell_type="EBV transformed B-lymphocytes"
     gene            1..1089
                     /gene="HLA-B"
                     /allele="B27052"
     CDS             1..1089
                     /gene="HLA-B"
                     /codon_start=1
                     /product="MHC class I antigen"
                     /protein_id="AAC42275.1"
                     /db_xref="GI:3643697"
                     /translation="MRVTEPRTLLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEP
                     RFITVGYVDDTLFVRFDSDAASPREEPRAPWIEQEGPEYWDRETQICKAKAQTDREDL
                     RTLLRYYNQSEAGSHTLQNMYGCDVGPDGRLLRGYHQDAYDGKDYIALNEDLSSWTAA
                     DTAAQITQRKWEAARVAEQLRAYLEGECVEWLRRYLENGKETLQRADPPKTHVTHHPI
                     SDHEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSG
                     EEQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRR
                     KSSGGKGGSYSQAACSDSAQGSDVSLTA"
BASE COUNT      217 a    332 c    368 g    172 t
ORIGIN
```

```
alignment_scores:
        Quality:    44.00              Length:       10
          Ratio:    4.889                Gaps:        0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x AF026218    .

Align seg 1/1  to: AF026218  from: 1  to: 1089

    1 ArgGluAspArgIleLeuLeuArgTyr 10
      ||||||||||||||||  ||||||||||
  295 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 324
```

OM of: US-08-653-294-14 to: N_Geneseq_36:*   out_format : pfs
Date: Feb 8, 2000 1:27 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-08-653-294-14
Query length: 10
Database: N_Geneseq_36:*
Database sequences 311585
Database length: 125096002
Search time (sec): 590.520000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
N_Geneseq_36:N70935	+	44.00	144.99	1.78	1026	! Sequence encoding the human hi
N_Geneseq_36:N70225	+	44.00	133.50	7.78	3874	! Sequence of genomic DNA encodi
N_Geneseq_36:T61639	+	44.00	128.95	13.94	6553	! HLA B27 consensus sequence. De
N_Geneseq_36:029167	+	39.00	138.25	4.23	270	! HLA-Bw 52 exon 2 alpha-1 domain
N_Geneseq_36:001834	+	39.00	126.20	19.83	1086	! Sequence encoding HLA-B51 anti
N_Geneseq_36:001822	+	39.00	126.20	19.83	1086	! Sequence encoding HLA-Bw52 ant
N_Geneseq_36:005693	+	39.00	126.18	19.89	1089	! HLA-B51 gene for production of
N_Geneseq_36:005701	+	39.00	126.18	19.89	1089	! HLA-Bw52 gene for production o
N_Geneseq_36:025121	+	39.00	126.18	19.89	1089	! HLA-Bw53 exon. HLA-B53 gene.
N_Geneseq_36:022114	+	39.00	126.18	19.89	1089	! HLA-Bw53 gene.
N_Geneseq_36:076405	+	34.00	122.00	34.00	213	! Human genome fragment. (Prefer
N_Geneseq_36:012083	-	34.00	120.96	38.82	240	! H.influenzae strain MinnA (OMP
N_Geneseq_36:X51732	-	34.00	108.80	184.62	978	! DNA encoding a human secreted p
N_Geneseq_36:T28520	-	34.00	104.55	318.39	1598	! H. influenzae detection probe
N_Geneseq_36:V74565	-	34.00	99.45	612.47	2881	! Staphylococcus aureus contig S
N_Geneseq_36:067406	-	34.00	98.76	669.83	3123	! Neural alpha-catenin protein c
N_Geneseq_36:V65390	-	34.00	98.42	699.42	3247	! Tomato ringspot virus peach is
N_Geneseq_36:V19871	-	34.00	93.57	1.3e+03	5688	! Rattus norvegicus cdo tumour s
N_Geneseq_36:V19008	+	34.00	93.05	1.3e+03	5688	! Rattus norvegicus Class II tum
N_Geneseq_36:V19007	+	34.00	93.06	1.4e+03	6030	! Rattus norvegicus Class II tum
N_Geneseq_36:V19870	-	34.00	93.05	1.4e+03	6039	! Rattus norvegicus cdo tumour s
N_Geneseq_36:X13170	+	34.00	88.22	2.6e+03	10555	! Enterococcus faecalis genome
N_Geneseq_36:V62176	+	34.00	67.38	3.6e+04	117213	! HSV-2 strain SB5 Contig ID 1
N_Geneseq_36:077006	+	33.00	119.37	47.62	189	! Human genome fragment. New nucl
N_Geneseq_36:047809	+	33.00	98.04	734.49	2223	! Aspartokinase II gene. DNA enc
N_Geneseq_36:T27730	+	33.00	97.89	748.08	2260	! Insulin-stimulated protein kin
N_Geneseq_36:T27731	+	33.00	97.89	748.08	2260	! Mutant insulin-stimulated prot
N_Geneseq_36:042561	+	33.00	95.56	1.0e+03	2960	! Histamine H1 receptor coding s
N_Geneseq_36:X00477	-	33.00	89.84	2.1e+03	5733	! Arabidopsis thaliana clavata1
N_Geneseq_36:X20262	-	33.00	85.60	3.6e+03	9359	! Borrelia burgdorferi polynucle
N_Geneseq_36:T11549	-	33.00	82.37	5.5e+03	13585	! Tumour rejection antigen prec
N_Geneseq_36:069946	+	32.00	125.78	20.91	59	! U7.6 L3' PCR primer for U7.6 var
N_Geneseq_36:T12619	+	32.00	123.15	29.32	80	! 2Cl1 scFv VL PCR primer 6. Singl
N_Geneseq_36:069941	+	32.00	122.83	30.55	83	! VK3 AL2 PCR primer for U7.6 vari
N_Geneseq_36:N92449	-	32.00	106.72	241.19	534	! Sequence of carcinoembryonic an
N_Geneseq_36:N50356	+	32.00	102.71	403.52	849	! Sequence encoding fused antibod
N_Geneseq_36:X13670	-	32.00	100.73	520.05	1067	! Enterococcus faecalis genome c
N_Geneseq_36:T01865	+	32.00	100.36	545.53	1114	! Fc(epsilon) CH2'-CH4 coding se
N_Geneseq_36:V21328	+	32.00	99.13	638.68	1284	! Human C epsilon exon. New immu
N_Geneseq_36:087474	+	32.00	98.97	651.95	1308	! Human IgE Fc chain (amino acid
N_Geneseq_36:091170	+	32.00	98.97	651.95	1308	! Human IgE Fc chain (amino acid
N_Geneseq_36:N40062	+	32.00	97.87	750.57	1485	! Sequence encoding human immund

N_Geneseq_36:N40063	+			32.00	96.92	847.67	1657	! Sequence of pGET2 encoding
N_Geneseq_36:071567	-			32.00	95.16	1.1e+03	2031	! Carcinoembryonic antigen DN
N_Geneseq_36:T36495	-			32.00	95.04	1.1e+03	2059	! Immunogenic carcinoembryon1
N_Geneseq_36:V80295	+			32.00	94.91	1.1e+03	2090	! scFv-IgG1 fusion construct

seq_name: N_Geneseq_36:N70935

seq_documentation_block:
ID   N70935 standard; DNA; 1026 BP.
AC   N70935;
DT   10-APR-1991 (first entry)
DE   Sequence encoding the human histocompatibility antigen HLA B27.
KW   Rheumatic disorder; genetic screening; diagnosis;
KW   ankylosing spondylitis; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   cds   1..1026
PN   DE3542024-A.
PD   04-JUN-1987.
PF   28-NOV-1985; 542024.
PR   28-NOV-1985; DE-542024.
PR   21-DEC-1985; DE-545576.
PA   (BEHW) BEHRINGWERKE AG.
PI   Riethmuller G, Meo T, Weiss E, Szots H;
DR   WPI; 87-157893/23.
DR   P-PSDB; P70590.
PT   DNA coding for antigen HLA B27 - and diagnostic reagents contg.
PT   such DNA, antigen or antibody
PS   Claim 2; Page 4; 5pp; German.
CC   The DNA may be used as a hybridisation probe for detecting the HLA
CC   B27 gene, e.g. for assessing susceptibility to rheumatic disorders
CC   such as ankylosis spondylitis, or may be used to transform cells
CC   for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27
CC   antibody in human serum, or to produce mono- or polyclonal HLA B27
CC   antibodies for use in immunoassay.
SQ   Sequence 1026 BP;   213 A;   307 C;   344 G;   162 T;

alignment_scores:
          Quality: 44.00   Length: 10
          Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x N70935

Align seg 1/1 to: N70935   from: 1 to: 1026

      1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
        ||||||||||||||| |||||||||||
    223 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 252

seq_name: N_Geneseq_36:N70225

seq_documentation_block:
ID   N70225 standard; DNA; 3874 BP.
AC   N70225;
DT   03-APR-1991 (first entry)
DE   Sequence of genomic DNA encoding human histocompatibility antigen
DE   HLA-B 27.
KW   Ankylosing spondylitis; rheumatic disorder; diagnosis; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   intron   518..590
FT   intron   720..989   /*tag= a
FT   intron   1090..1506   /*tag= b
FT   intron   1932..2357   /*tag= c
FT   intron   2450..2566   /*tag= d
FT   intron   /*tag= e

```
FT   intron           3009..3041
FT                    /*tag= f
FT   intron           3148..3191
FT                    /*tag= g
PN   EP-226069-A.
PD   24-JUN-1987.
PF   21-NOV-1986; 116139.
PR   01-JAN-1985; DE-542024.
PR   21-DEC-1985; DE-545576.
PA   (BEHW ) BEHRINGWERKE AG.
PI   Szots H, Weiss E, Dorner C, Lang M, Meo T, Riethmuller G;
PR   WPI; 87-171469/25.
DR   P-PSDB; P70155.
PT   DNA coding for human histocompatibility antigen HLA-B 27 - useful
PT   for diagnosis and antigen and antibody prodn.
PS   Claim 1; p6; 13pp; German.
CC   THe DNA may be used to detect the HLA-B 27 gene (opt. mutated) in
CC   human genetic material.  The HLA-B 27 may be used to detect anti-HLA-
CC   B 27 antibodies in human serum.  The antibodies may be used to
CC   determine HLA-B 27 levels in human serum, e.g. for diagnosis of
CC   rheumatic disorders, esp. ankylosing spondylitis.
SQ   Sequence 3874 BP;  751 A;  1094 C;  1171 G;  858 T;

alignment_scores:
         Quality:    44.00          Length:      10
         Ratio:    4.889           Gaps:       0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x N70225   ..

Align seg 1/1  to: N70225  from: 1  to: 3874

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||||||||||  |||||||||||
  941 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 970

seq_name: N_Geneseq_36:T61639

seq_documentation_block:
ID   T61639 standard; DNA; 6553 BP.
AC   T61639;
DT   05-JUN-1997 (first entry)
DE   HLA B27 consensus sequence.
KW   HLA B27; seronegative spondylarthropathy; ankylosing spondylitis;
KW   Reiter's syndrome; arthritis; acute anterior uveitis; diagnosis;
KW   ss; ds.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   mrna           3968..6653
FT                  /*tag= a
FT                  /note= "HLA-B27 3' flanking region, downstream of
FT                  3' untranslated region"
FT   mrna           4112..4556
FT                  /*tag= b
FT                  /note= "3' flanking region diagnostic for genetic
FT                  predisposition to SNSA"
FT   mrna           4270..4556
FT                  /*tag= b
FT                  /note= "3' flanking region diagnostic for genetic
FT                  predisposition to SNSA"
FT   misc_difference 4495
FT                  /*tag= d
FT                  /note= "absence of cytosine at this site is
FT                  indicative of a predisposition to SNSA"
PN   WO9709450-A1.
PD   13-MAR-1997.
PF   16-AUG-1996; U13256.
PR   01-SEP-1995; US-522942.
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
PI   Tyan DB;
DR   WPI; 97-192924/17.


PT   Detecting pre-disposition to seronegative spondylarthropathies -
PT   from the absence of a C residue at a specific position in the
PT   3'-flanking region of the HLA B27 allele
PS   Claim 1; Page 52-56; 68pp; English.
CC   Genetic predisposition to seronegative spondyloarthropathies (SNSA)
CC   is detected by determining the absence of a cytosine nucleotide in
CC   the 3' flanking region (see also T61647-48) of an HLA-B gene at a
CC   position corresponding to nucleotide 4495 of the HLA-B27 consensus
CC   sequence given in T61639.  Probes and primers (see also T61640-46)
CC   based on this region can be used in diagnostic assays to detect the
CC   genetic predisposition to SNSA, and permit the distinction of B27+
CC   individuals who are resistant to SNSA from B27+ normal individuals
CC   who are susceptible (but as yet unaffected) to such diseases.
SQ   Sequence 6553 BP;  1443 A;  1619 C;  2017 G;  1474 T;

alignment_scores:
         Quality:    44.00          Length:      10
         Ratio:    4.889           Gaps:       0
Percent Similarity:  90.000    Percent Identity:  90.000

alignment_block:
US-08-653-294-14 x T61639   ..

Align seg 1/1  to: T61639  from: 1  to: 6553

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||||||||||||||  |||||||||||
 1102 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 1131

seq_name: N_Geneseq_36:Q29167

seq_documentation_block:
ID   Q29167 standard; DNA; 270 BP.
AC   Q29167;
DT   09-MAR-1993 (first entry)
DE   HLA-Bw 52 exon 2 alpha-1 domain.
KW   Human leukocyte antigen; transgenic; germ cells; somatic cells;
KW   expression; ss.
PN   J04091731-A.
PD   25-MAR-1992.
PF   03-AUG-1990; 207329.
PR   03-AUG-1990; JP-207329.
PA   (OLYU ) OLYMPUS OPTICAL CO.
DR   WPI; 92-342893/42.
PT   Transgenic non-human mammalian HLA-Bw 52 gene - useful for
PT   analysis of expression of gene structure, and prodn. of
PT   mouse model of human disease
PS   Disclosure; Fig 1; 8pp; Japanese.
CC   The sequence shows the exon 2 alpha-1-domain of the human leukocyte
CC   antigen-Bw 52 gene.  The complete gene may be introduced into non-
CC   human mammals, pref. rat or mouse, or their ancestors at the primary
CC   developmental biological step via transplantation into the zygote or
CC   embryo to generate transgenic non-human mammals incorporating the
CC   HLA-Bw 52 gene in both their germ cells and somatic cells. Transgenic
CC   non-human mammals contg. HLA-Bw 52 are useful for the analysis of
CC   expression of the gene, its structure, and prodn. of mouse models of
CC   human disease.           See also Q29166-72.
SQ   Sequence  270 BP;    59 A;     88 C;    86 G;     37 T;

alignment_scores:
         Quality:    39.00          Length:      10
         Ratio:    4.333           Gaps:       0
Percent Similarity:  90.000    Percent Identity:  80.000

alignment_block:
US-08-653-294-14 x Q29167   ..

Align seg 1/1  to: Q29167  from: 1  to: 270

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      ||||||::::|||||||||  |||||||||
```

```
     222 CGAGAGAACCTGCGGATCGGCGTCCGCTAC 251

seq_name: N_Geneseq_36:Q01834

seq_documentation_block:
ID   Q01834 standard; DNA; 1086 BP.
AC   Q01834;
DT   19-MAR-1991 (first entry)
DE   Sequence encoding HLA-B51 antigen.
KW   Probe; HLA class I DNA; immunogen; ss.
OS   Homo sapiens.
PN   EP354580-A.
PD   14-FEB-1990.
PF   10-AUG-1989.
PR   11-AUG-1988; JP-200758.
PA   (OLYU) Olympus Optical Co., Ltd.
PI   Kano K, Takiguchi;
DR   WPI; 90-046289/07.
PT   New DNA for class 1 human leucocyte antigens and derived probes and
PT   transformed cells, useful for DNA typing, as immunogens etc.
PS   Claim 1; Page 11; 23pp; English.
CC   The HLA class I DNA can be used a source of probes for use in DNA
CC   typing. Transformed cells, which are useful as immunogens, can be
CC   obtained by introducing these DNAs into eucaryotic cells.
SQ   Sequence 1086 BP;    224 A;    334 C;    356 G;    172 T;

alignment_scores:
       Quality:    39.00              Length:      10
       Ratio:      4.333              Gaps:        0
Percent Similarity: 90.000      Percent Identity: 80.000

alignment_block:
US-08-653-294-14 x Q01834

Align seg 1/1 to: Q01834 from: 1 to: 1086

       1 ArgGluAspLeuArgIleLeuArgTyr 10
         |||||::||||||||  ||||||||
     294 CGAGAGAACCTGCGGATCGGCGTCCGCTAC 323

seq_name: N_Geneseq_36:Q01822

seq_documentation_block:
ID   Q01822 standard; DNA; 1086 BP.
AC   Q01822;
DT   19-MAY-1991 (first entry)
DE   Sequence encoding HLA-Bw52 antigen.
KW   Probe; HLA class I DNA; immunogen; ss.
OS   Homo sapiens.
FH   Key       Location/Qualifiers
FT   cds       1..1086
FT             /*tag= a
PN   EP-354580-A.
PD   14-FEB-1990.
PF   10-AUG-1989.
PR   11-AUG-1988; JP-200758.
PA   (OLYU) Olympus Optical Co., Ltd.
PI   Kano K, Takiguchi;
WPI  90-046289/07.
DR   P-PSDB; R03142.
PT   New DNA for class 1 human leucocyte antigens and derived probes and
PT   transformed cells, useful for DNA typing, as immunogens etc.
PS   Claim 2; pp11-12; 23pp; English.
CC   The HLA class I DNA can be used a source of probes for use in DNA
CC   typing. Transformed cells, which are useful as immunogens, can be
CC   obtained by introducing these DNAs into eucaryotic cells.
SQ   Sequence 1086 BP;    223 A;    335 C;    358 G;    170 T;

alignment_scores:
       Quality:    39.00              Length:      10
       Ratio:      4.333              Gaps:        0
```

```
Percent Similarity: 90.000      Percent Identity: 80.000

alignment_block:
US-08-653-294-14 x Q01822

Align seg 1/1 to: Q01822 from: 1 to: 1086

       1 ArgGluAspLeuArgIleLeuArgTyr 10
         |||||::||||||||  ||||||||
     294 CGAGAGAACCTGCGGATCGGCGTCCGCTAC 323

seq_name: N_Geneseq_36:Q05693

seq_documentation_block:
ID   Q05693 standard; DNA; 1089 BP.
AC   Q05693;
DT   03-JAN-1991 (first entry)
DE   HLA-B51 gene for production of monoclonal antibodies.
KW   Alloptype specific monoclonal anti-HLA antibodies; hybridomas;
KW   transgenic animals; HLA-B51 gene; ss.
FH   Key       Location/Qualifiers
FT   exon      1..73
FT             /*tag= a
FT             /number=1
FT   exon      74..343
FT             /*tag= b
FT             /number=2
FT             /note="alpha 1-domain"
FT   exon      344..619
FT             /*tag= c
FT             /number=3
FT             /note="alpha 2-domain"
FT   intron    620..895
FT             /*tag= d
FT             /number=4
FT   exon      896..1012
FT             /*tag= e
FT             /number=5
FT             /note="alpha 3-domain"
FT   exon      1013..1042
FT             /*tag= f
FT             /number=6
FT   exon      1043..1089
FT             /*tag= g
FT             /number=7
PN   EP-383183-A.
PD   22-AUG-1990.
PF   07-FEB-1990; 102424.
PR   08-FEB-1989; JP-029313.
PA   (OLYU ) OLYMPUS OPTICAL KK.
PI   Takiguchi M;
DR   WPI; 90-255479/34.
PT   Allotype specific monoclonal anti- HLA antibodies prodn. - using
PT   hybridomas derived from transgenic animals carrying HLA gene and
PT   immunised with HLA antigen of different allotype
PS   Disclosure; Fig 1 A-G; 20pp; English.
CC   The human HLA-B51 gene was injected into fertilised mouse eggs and
CC   then these introduced into the uterus of a pseudo pregnant mouse.
CC   The young were tested to ensure incorporation of the gene into the
CC   chromosome, and one of them mated 3 times with a normal male to
CC   produce 16 young, seven of which carried the HLA-B51 gene.
CC   The transgeneic offspring were immunised with HLA antigen.
CC   The spleen lymphocytes were fused with myeloma cells. Hybridomas
CC   producing antibodies were selected.
CC   See also Q05701.
SQ   Sequence 1089 BP;    224 A;    335 C;    357 G;    173 T;

alignment_scores:
       Quality:    39.00              Length:      10
       Ratio:      4.333              Gaps:        0
Percent Similarity: 90.000      Percent Identity: 80.000
```

```
alignment_block:
US-08-653-294-14 x Q05693

Align seq 1/1  to: Q05693  from: 1  to: 1089

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      |||||||:::|||||||||  |||||||||
  295 CGAGAGAACCTGCGGATCGCGGCTCCGCTAC 324

seq_name: N_Geneseq_36:Q05701

seq_documentation_block:
ID  Q05701 standard; DNA; 1089 BP.
AC  Q05701;
DT  03-JAN-1991 (first entry)
DE  HLA-Bw52 gene for production of monoclonal antibodies.
KW  Alloptype specific monoclonal anti-HLA antibodies; hybridomas;
KW  transgenic animals; HLA-Bw52 gene; ss.
FH  Key             Location/Qualifiers
FT  exon            1..73
FT                  /*tag= a
FT                  /number=1
FT  exon            74..343
FT                  /*tag= b
FT                  /number=2
FT                  /note="alpha 1-domain"
FT  exon            344..619
FT                  /*tag= c
FT                  /number=3
FT                  /note="alpha 2-domain"
FT  intron          620..895
FT                  /*tag= d
FT                  /number=4
FT                  /note="alpha 3-domain"
FT  exon            896..1012
FT                  /*tag= e
FT                  /number=5
FT  exon            1013..1042
FT                  /*tag= f
FT                  /number=6
FT  exon            1043..1089
FT                  /*tag= g
FT                  /number=7
PN  EP-383183-A.
PD  22-AUG-1990.
PF  07-FEB-1990; 102424.
PR  08-FEB-1989; JP-029313.
PA  (OLYU ) OLYMPUS OPTICAL KK.
PI  Takiguchi M;
DR  WPI; 90-255479/34.
PT  Allotype specific monoclonal anti- HLA antibodies prodn. - using
PT  hybridomas derived from transgenic animals carrying HLA gene and
PT  immunised with HLA antigen of different allotype
PS  Disclosure: Fig 1 A-G; 20pp; English.
CC  The human HLA-Bw52 gene was introduced into mouse L cells and
CC  then these cells used to immunise one of the transgenic mice
CC  (See Q05693).
CC  The spleen lymphocytes were fused with myeloma cells (P3x63-Ag8.653).
CC  Hybridomas producing antibodies were selected.
SQ  Sequence  1089 BP;  223 A;  336 C;  359 G;  171 T;

alignment_scores:
        Quality:     39.00        Length:    10
          Ratio:      4.333         Gaps:      0
Percent Similarity:  90.000   Percent Identity:  80.000

alignment_block:
US-08-653-294-14 x Q05701

Align seq 1/1  to: Q05701  from: 1  to: 1089

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
```

```
      |||||||:::|||||||||  |||||||||
  295 CGAGAGAACCTGCGGATCGCGGCTCCGCTAC 324

seq_name: N_Geneseq_36:Q12114

seq_documentation_block:
ID  Q12114 standard; DNA; 1089 BP.
AC  Q12114;
DT  29-AUG-1991 (first entry)
DE  HLA-Bw53 exon.
KW  Human leukocyte antigen; probe; major histocompatibility complex;
KW  MHC; class I; ss.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  cds             1..1089
FT                  /*tag= a
PN  J03112487-A.
PD  14-MAY-1991.
PF  22-SEP-1989; 247697.
PR  22-SEP-1989; JP-247697.
PA  (OLYU ) OLYMPUS OPTICAL KK.
DR  WPI; 91-182991/25.
DR  P-PSDB; R12463.
PT  HLA-Bw53 gene, DNA probe and transformant cells - used for
PT  immunisation, identifying specificity of antiserum etc.
PS  Claim 1; Page 1; 11pp; Japanese.
CC  Probes comprising part of the sequence can be used to identify
CC  Class I genes. The DNA can be expressed for immunisation of
CC  animals and prodn. of monoclonal antibodies specific for the
CC  HLA-Bw53 antigen.  See also J03112485 and J03112486.
SQ  Sequence  1089 BP;  222 A;  337 C;  356 G;  174 T;

alignment_scores:
        Quality:     39.00        Length:    10
          Ratio:      4.333         Gaps:      0
Percent Similarity:  90.000   Percent Identity:  80.000

alignment_block:
US-08-653-294-14 x Q12114

Align seq 1/1  to: Q12114  from: 1  to: 1089

    1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
      |||||||:::|||||||||  |||||||||
  295 CGAGAGAACCTGCGGATCGCGGCTCCGCTAC 324

seq_name: N_Geneseq_36:Q76405

seq_documentation_block:
ID  Q76405 standard; DNA; 213 BP.
AC  Q76405;
DT  23-SEP-1994 (first entry)
DE  Human genome fragment. (Preferred)
KW  Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW  detection; homology; human; adrenal tissue; ds.
OS  Homo sapiens.
PN  WO9401548-A.
PD  20-JAN-1994.
PF  13-JUL-1993; G01467.
PR  13-JUL-1992; GB-014857.
PA  (MEDI-) MEDICAL RES COUNCIL.
PI  Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI  Sibson DR, Starkey M;
DR  WPI; 94-035056/04.
PT  New nucleic acid fragment encoding gene products - can be used
PT  for genetic analysis and mapping
PS  Claim 1; Page 23-24; 616pp; English.
CC  Human nucleic acid fragments, isolated from brain adrenal tissue,
CC  the placenta or bone marrow comprise any of: (A) a sequence
CC  selected from (Q76401-Q77613), (B) an allelic variation of a
CC  sequence as described in (A), or (C) a sequence complementary
CC  to (A) or (B).
```

```
CC  Preferred sequences exhibit no more than 90% homology to a human
CC  sequence known per se.
SQ  Sequence 213 BP;   76 A;   39 C;   41 G;   57 T;

alignment_scores:
        Quality:   34.00          Length:        9
        Ratio:      4.250          Gaps:          0
Percent Similarity: 88.889     Percent Identity: 77.778

alignment_block:
US-08-653-294-14 x Q76405      ..

Align seg 1/1 to: Q76405  from: 1  to: 213

     1 ArgGluAspLeuArgIleLeuLeuArg 9
       |||||||||||||||||:::: |||
   160 AGGGAGGATTTAAGAATAATGAAAAGA 186

seq_name: N_Geneseq_36:Q12083

seq_documentation_block:
ID  Q12083 standard; DNA; 240 BP.
AC  Q12083;
DT  26-JUL-1991 (first entry)
DE  H.influenzae strain MinnA (OMP subtype 1H)-cro-lacz fusion
DE  construct.
KW  bacterial meningitis; vaccine; P1 gene; T-cell antigen; pRSM793;
KW  outer membrane protein; ss.
OS  Haemophilus influenzae.
FH  Key           Location/Qualifiers
FT  cds           1..234
FT                /tag= a
FT                /product= partial cro-lacz-strain MinnA P1 OMP
FT                fusion protein
PN  WO9106552-A.
PD  16-MAY-1991.
PF  31-OCT-1990; CA0374.
PR  31-OCT-1989; GB-024473.
PA  (CONN-) CONNAUGHT LAB LTD.
PA  (UNIW ) WASHINGTON UNIV ST. LOUIS.
PI  Munson RS, Grass S, Chong P, Yang Y, Fahim R, McVerry P;
PI  Klein M;
DR  WPI; 91-164201/22.
PT  Outer membrane protein of Haemophilus influenzae type B - used as
PT  vaccine against infections, esp. in infants and for diagnosis
PS  Disclosure; Fig 5; 33pp; English.
CC  Plasmid pRSM793 contains only the 3' portion of the P1 gene. The
CC  plasmid is derived from pRSM188 which contains the full-length P1
CC  sequence. The cro-lacz-ompP1 fusion protein produced from pRSM793
CC  was recognised by rabbit and guinea pig P1-specific antisera in
CC  immunoblot analyses.
CC  See also R12446-R12455.
SQ  Sequence 240 BP;   83 A;   44 C;   38 G;   75 T;

alignment_scores:
        Quality:   34.00          Length:       10
        Ratio:      3.778          Gaps:          0
Percent Similarity: 90.000     Percent Identity: 60.000

alignment_block:
US-08-653-294-14 x Q12083/rev      ..

Align seg 1/1 to reverse of: Q12083  from: 1  to: 240

     1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
       ||||::||||||||||:::||||
   208 CGTAAAGATTTGCGTGCGTGCTTGAGAAGTAT 179

seq_name: N_Geneseq_36:X51732

seq_documentation_block:


ID  X51732 standard; DNA; 978 BP.
AC  X51732;
DT  17-JUN-1999 (first entry)
DE  DNA encoding a human secreted protein.
KW  Human secreted protein; cancer; immune disorder; infection:
KW  inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW  restenosis; autoimmune disorder; Alzheimer's disease;
KW  peripheral neuropathy; trauma; spinal cord injury; allergy;
KW  hematopoietic disorder; skeletal disorder; neurological disorder;
KW  arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW  transplant rejection; ss.
OS  Homo sapiens.
PN  WO9911293-A1.
PD  11-MAR-1999.
PF  03-SEP-1998; U18360.
PR  12-SEP-1997; US-058974.
PR  05-SEP-1997; US-057626.
PR  05-SEP-1997; US-057663.
PR  05-SEP-1997; US-057669.
PR  12-SEP-1997; US-058666.
PR  12-SEP-1997; US-058667.
PR  12-SEP-1997; US-058973.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI  Rosen GA, Ruben SM, Shi Y;
DR  WPI; 99-204988/17.
DR  P-PSDB; Y12945.
PT  New isolated human genes and the secreted polypeptides they encode
PT  - useful for diagnosis and treatment of e.g. neurological disorders,
PT  tumours, immune disorders, inflammation or haematological disorders
PS  Claim 1; Page 170; 215pp; English.
CC  X51701-55 encode human secreted proteins. The polynucleotides and
CC  their corresponding secreted polypeptides are useful for preventing,
CC  treating or ameliorating medical conditions, e.g. by protein or gene
CC  therapy. Pathological conditions can also be diagnosed by determining
CC  the amount of the new polypeptides in a sample or by determining the
CC  presence of mutations in the new polynucleotides. Specific uses are
CC  described for each polynucleotide, based on which tissues they are
CC  most highly expressed in, and include developing products for the
CC  diagnosis or treatment of cancer, immune disorders, infection,
CC  inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC  restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC  neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC  disorders, skeletal disorders, neurological disorders, arthritic
CC  disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC  rejection. The polypeptides are also useful for identifying their
CC  binding partners.
SQ  Sequence 978 BP;   285 A;   256 C;   197 G;   240 T;

alignment_scores:
        Quality:   34.00          Length:       10
        Ratio:      3.778          Gaps:          0
Percent Similarity: 90.000     Percent Identity: 70.000

alignment_block:
US-08-653-294-14 x X51732/rev      ..

Align seg 1/1 to reverse of: X51732  from: 1  to: 978

     1 ArgGluAspLeuArgIleLeuArgTyr 10
       ||||::|||||||||| ||||||:::
    99 AGGGAGGAGCTGAGAATCCAACTGCGGTGG 70

seq_name: N_Geneseq_36:T28520

seq_documentation_block:
ID  T28520 standard; DNA; 1598 BP.
AC  T28520;
DT  02-APR-1997 (first entry)
DE  H. influenzae detection probe #2.
KW  Detection; probe; amplification primer; bacterial pathogen; pneumonia;
KW  Escherichia coli; Klebsiella pneumoniae; Pseudomonas aeruginosa;
```

KW   Proteus mirabilis; Streptococcus pneumoniae; Staphylococcus aureus;
KW   Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;
KW   Staphylococcus saprophyticus; Streptococcus pyogenes; urinary tract;
KW   Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;
KW   infection; intra-abdominal infection; skin infection;
KW   bacterial resistance; beta-lactam antibiotic; ds.
OS   Synthetic.
PN   WO9600582-A2.
PD   21-MAR-1996.
PF   12-SEP-1995; CA0528.
PR   12-SEP-1994; US-304732.
PA   (BERG/) BERGERON M G.
PA   (OUEL/) OUELLETTE M.
PA   (ROYP/) ROY P H.
PI   Bergeron MG, Ouellette M, Roy PH;
DR   WPI; 96-179953/18.
PT   Method for the detection of bacterial species using probes and
PT   primers - allows detection and quantification of antibiotic
PT   resistant bacteria in patients, the environment and food
PS   Claim 47; Page 80-81; 216pp; English.
CC   The sequences given in T28519-24 represent probes which were used in
CC   the method of the invention for the detection of H. influenzae in a
CC   sample. The method comprises using probes and/or amplification primers
CC   which are specific, ubiquitous and sensitive for determining the presence
CC   and/or amount of nucleic acids from selected bacterial species in any
CC   sample, where the bacterial nucleic acid comprises a selected target
CC   region hybridisable with the probes or primers. The method comprises
CC   contacting the sample with the probes or primers and detecting the
CC   presence and/or amount of hybridised primers or amplification products
CC   as and indication of the presence and/or amount of the bacterial
CC   species. This method may be used to detect commonly encountered
CC   bacterial pathogens, e.g. Escherichia coli, Klebsiella pneumoniae,
CC   Pseudomonas aeruginosa, Proteus mirabilis, Streptococcus pneumoniae,
CC   Staphylococcus aureus, Staphylococcus epidermidis, Enterococcus
CC   faecalis, Staphylococcus saprophyticus, Streptococcus pyogenes,
CC   Haemophilus influenzae and Moraxella catarrhalis. These bacterial
CC   species are associated with approx. 90% of urinary tract infections and
CC   with a high percentage of other severe infections including
CC   septicaemia, meningitis, pneumonia, intra-abdominal infections, skin
CC   infections and other severe respiratory tract infections. The method
CC   may also be used to evaluate a bacterial resistance to beta-lactam
CC   antibiotics.
SQ   Sequence 1598 BP; 500 A; 278 C; 329 G; 491 T;

alignment_scores:
      Quality: 34.00   Length: 10
      Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-14 x T28520/rev

Align seg 1/1 to reverse of: T28520 from: 1 to: 1598

     1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
       ||::||:||||||||:::||| ::||:|||
  1425 CGTAAAGATTTGCGTGCTTGAGAAGTAT 1396

seq_name: N_Geneseq_36:V74565

seq_documentation_block:
ID   V74565 standard: DNA: 2881 BP.
AC   V74565;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #254.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   misc_feature   541 .600
FT   /tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   2341 .2400
FT   /tag= b
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1047-1049; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 2881 BP; 1049 A; 376 C; 505 G; 826 T;

alignment_scores:
      Quality: 34.00   Length: 9
      Ratio: 3.778   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
US-08-653-294-14 x V74565/rev

Align seg 1/1 to reverse of: V74565 from: 1 to: 2881

     2 GluAspLeuArgIleLeuLeuArgTyr 10
       :::||||||:::||||||||||:::
   948 GACGATTAAAATTCTGCTAAGATTT 922

seq_name: N_Geneseq_36:Q67406

seq_documentation_block:
ID   Q67406 standard; cDNA to mRNA; 3123 BP.
AC   Q67406;
DT   19-APR-1995 (first entry)
DE   Neural alpha-catenin protein coding sequence.
KW   Adhesion; neural alpha catenin; tumour; metastasis; disease;
KW   autoimmune disease; neural alpha catenin; infectious disease; dermal disease;
KW   arteriosclerosis; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   cds   125 .2845
FT   /tag= a
FT   /product= Neural alpha catenin.

```
PN   J06211898-A.
PD   02-AUG-1994.
PF   25-DEC-1992; 358026.
PR   25-DEC-1992; JP-358026.
PA   (TAKI ) TAKARA SHUZO CO LTD.
DR   WPI: 94-283359/35.
DR   P-PSDB: R58778.
PT   Neural alpha-catenin protein and DNA - useful in the control of
PT   cell adhesion, e.g., in treatment of tumour (metastasis) and
PT   autoimmune disease
PS   Disclosure; Page 10-14; 14pp; Japanese.
CC   The neural alpha catenin can be used for the treatmnet of diseases
CC   related to intercellular adhesion such as primary tumour, tumour
CC   metastasis, autoimmune diseases, infectious diseases, dermal
CC   diseases and arteriosclerosis.
SQ   Sequence  3123 BP;   920 A;   674 C;   827 G;   702 T;
```

```
alignment_scores:
         Quality:    34.00        Length:    9
         Ratio:     3.778           Gaps:    0
Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
US-08-653-294-14 x Q67406/rev   ..

Align seg 1/1  to reverse of: Q67406  from: 1  to: 3123

       2 GluAspLeuArgIleLeuLeuArgTyr 10
         |||//||||||::::||||||:::::::
    1390 GAAGACCTGCGGCGTACTCCTTCACTTC 1364
```

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-14 to: EST:*   out_format : pfs

Date: Feb 8, 2000 4:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160700_15770/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-14
Query length: 10
Database: EST:*
Database sequences: 4538634
Database length: 188783982
Search time (sec): 8553.360000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_est38:AL036690	+	44.00	165.79	1.81	171	; AL036690 DKF2p564D2463_r1 564 (
gb_est30:AU056838	+	41.00	142.61	35.40	701	; AU056838 AU056838 Oryza sativa
gb_est21:AA989542	-	40.00	143.34	32.21	402	! AA989542 am64d02.s1 Barstead sp
gb_gss13:AQ440876	-	40.00	141.53	40.65	501	! AQ440876 HS_5098_B2_B04_T7A RPC
gb_est21:AA975627	+	40.00	141.46	41.00	505	! AA975627 oq63b05.s1 NCI_CGAP_Ki
gb_est8:C03945	+	39.00	144.03	29.51	232	! C03945 C03945 Human heart cDNA
gb_est10:AA151891	+	39.00	143.25	32.61	255	! AA151891 zo01f06.r1 Stratagene
gb_est21:AA952680	+	39.00	142.81	34.50	269	! AA952680 TENS1864 T. cruzi epim
gb_est11:AA263158	+	39.00	142.39	36.41	283	! AA263158 PMY0534 KG1-a Lambda Z
gb_est6:D82221	+	39.00	140.07	49.03	375	! D82221 HUMHBC4626 Human pancrea
gb_gss13:AQ444169	-	39.00	139.51	52.63	401	! AQ444169 GSSTc0231 Trypanosoma
gb_est37:AI957215	-	39.00	139.05	55.83	424	! AI957215 u177a10.x1 Sugano mous
gb_est10:AA147151	-	39.00	136.46	77.89	581	! AA147151 zo32d06.r1 Stratagene
gb_est26:AI359260	-	39.00	135.95	83.15	618	! AI359260 qy27b07.x1 NCI_CGAP_Br
gb_gss13:AQ449604	+	39.00	134.75	97.01	715	! AQ449604 500002D08.x2 CPIOWAM13
gb_est31:AI696864	+	39.00	134.37	101.75	748	! AI696864 wc74h11.x1 NCI_CGAP_Pa
gb_gss8:AQ039738	-	38.00	136.11	81.45	380	! AQ039738 CIT-HSP-2317E17.TF CIT
gb_est37:AI946856	+	38.00	135.28	90.55	420	! AI946856 bs31h08.y1 Drosophila
gb_gss13:AQ440598	-	38.00	135.09	92.83	430	! AQ440598 HS_5089_B1_C10.SP6E RP
gb_est22:AI028215	+	38.00	135.01	93.74	434	! AI028215 ov96c07.x1 Soares_test
gb_est11:AQ301014	+	38.00	134.61	98.77	456	! AQ301014 HS_3105_A2_E01_MR CIT
gb_est9:AA082472	-	38.00	134.46	100.61	464	! AA082472 zn40a08.r1 Stratagene
gb_gss3:B61151	+	38.00	134.18	104.28	480	! B61151 T2005TF TAMU Arabidopsis
gb_est9:AA082478	-	38.00	132.72	125.75	573	! AA082478 zn40b07.r1 Stratagene
gb_est37:AI946939	+	38.00	132.47	129.94	591	! AI946939 bs33c03.y1 Drosophila
gb_est24:AI239094	+	38.00	131.90	139.72	633	! AI239094 GH15272.5prime GH Dros
gb_est8:AQ080015	+	37.00	133.78	109.80	316	! AQ080015 CIT-HSP-2367F12.TR CIT
gb_est20:AA879637	+	37.00	133.01	121.22	347	! AA879637 mt68h02.r1 Soares mous
gb_est10:AA015279	+	37.00	132.50	129.36	369	! AA015279 mh33d03.r1 Soares mous
gb_est32:AI748514	-	37.00	131.80	141.62	402	! AI748514 sb54a12.y1 Gm-c1016 Gl
gb_est28:AI508196	+	37.00	131.49	147.22	417	! AI508196 mh33d03.y1 Soares mous
gb_est11:AA239196	-	37.00	131.41	148.71	421	! AA239196 mx89c04.r1 Soares mous
gb_gss3:B46971	+	37.00	131.14	153.94	435	! B46971 HS-1066-A2-E07-MR.abi CI
gb_est9:AQ435812	+	37.00	130.96	157.69	445	! AQ435812 HS_5063_A1_B10.SP6E RP
gb_est16:AA596937	+	37.00	130.92	158.44	447	! AA596937 vo14h04.r1 Barstead mc
gb_est16:AA592217	-	37.00	130.76	161.81	456	! AA592217 vo1se03.r1 Barstead mc
gb_gss10:AQ199154	-	37.00	130.54	166.32	468	! AQ199154 RPCI11-45E5.TK RPCI-11
gb_gss12:AQ370195	-	37.00	129.98	178.75	501	! AQ370195 HS_5045_B1_E03_T7 RPCI
gb_est12:AU014732	+	37.00	129.75	184.03	515	! AU014732 Mouse two-cel
gb_est25:AU045014	+	37.00	129.26	196.15	547	! AU045014 AU045014 Mouse sixteen
gb_est39:AW119564	+	37.00	128.82	207.54	577	! AW119564 sd48a12.y1 Gm-c1016 Gl
		37.00	128.63	212.49	590	

gb_est22:AI055656   +   37.00   127.97   231.19   639   ! AI055656 coau0004K01 Cotton
gb_est44:AW208428   +   37.00   127.29   252.28   694   ! AW208428 uo60c03.x1 NCI_CGAP
gb_gss3:B20285   +   37.00   126.40   282.75   773   ! B20285 T2OU7-T7 TAMU Arabido

seq_name: gb_est38:AL036690

seq_documentation_block:
LOCUS   AL036690   171 bp   mRNA   EST   27-SEP-1999
DEFINITION DKF2p564D2463_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
            DKF2p564D2463 5', mRNA sequence.
ACCESSION   AL036690
VERSION   AL036690.3 GI:5927859
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 171)
AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
TITLE   EST (Duesterhoeft, et al.)
JOURNAL   Unpublished (1999)
COMMENT   On Jul 7, 1999 this sequence version replaced gi:5866258.
            Contact: Duesterhoeft A
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen within the cDNA sequencing consortium of the
            German Genome Project.
            No sequence available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. .
FEATURES   Location/Qualifiers
     source   1..171
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKF2p564D2463"
                     /clone_lib="564 (synonym: hfbr2)"
                     /tissue_type="brain"
                     /dev_stage="fetal"
                     /lab_host="xl-2blue"
                     /note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT   36 a   53 c   60 g   22 t
ORIGIN

alignment_scores:
                Quality: 44.00   Length: 10
                Ratio: 4.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-14   from: 1 to: 171

Align seg 1/1 to: AL036690 from: 1 to: 171

          1 ArgGluAspLeuArgIleLeuArgTyr 10
            !!!!!!!!!!!!!!| !!!!!!!!!!!!!
         62 CGAGAGGACCTGCGGACCCTGCTCCGCTAC 91

seq_name: gb_est30:AU056838

seq_documentation_block:
LOCUS   AU056838   701 bp   mRNA   EST   29-APR-1999
DEFINITION AU056838 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
            clone S20019_1A, mRNA sequence.
ACCESSION   AU056838
VERSION   AU056838.1 GI:4715722
KEYWORDS   EST.
SOURCE   Oryza sativa.

ORGANISM   Oryza sativa
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
           Poaceae; Oryza.
REFERENCE 1 (bases 1 to 701)
AUTHORS   Yamamoto,K. and Sasaki,T.
TITLE   Rice cDNA from mature leaf
JOURNAL   Unpublished (1999)
COMMENT   On Jun 5, 1998 this sequence version replaced gi:3187083.
           Contact: Takuji Sasaki
           National Institute of Agrobiological Resources
           Rice Genome Research Program
           2-1-2 Kannondai,Tsukuba
           Ibaraki,
           Japan 305
           Tel: 0298-38-7441
           Fax: 0298-38-7468
           Email: tasaki@abr.affrc.go.jp
           PROJECT ='RGP'
FEATURES   Location/Qualifiers
     source   1..701
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S20919_1A"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT   145 a   169 g   230 g   151 t   6 others
ORIGIN

alignment_scores:
          Quality: 41.00   Length: 10
          Ratio: 4.556   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-14 x AU056838/rev   ..

Align seg 1/1 to reverse of: AU056838 from: 1 to: 701

      1 ArgGluAspArgIleLeuLeuArgTyr 10
        ||| |||||||||||||||||||||||
     92 AGGCTTGATCTCAGAATTCTCTTCAGATAC 63

seq_name: gb_est21:AA989542

seq_documentation_block:
LOCUS   AA989542   402 bp   mRNA   EST   02-JUN-1998
DEFINITION am64d02.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
            IMAGE:1576803 3' similar to gb:L05093 60S RIBOSOMAL PROTEIN L18A
            (HUMAN), mRNA sequence.
ACCESSION   AA989542
VERSION   AA989542.1 GI:3174906
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 402)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE   WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2153091.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1576803"
                     /clone_lib="Barstead spleen HPLRB2"
                     /sex="male"
                     /dev_stage="adult, 17 years"
                     /lab_host="DH10B"
                     /note="Organ: spleen; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCGGCGCCCTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. "
BASE COUNT   72 a   106 c   132 g   92 t
ORIGIN

alignment_scores:
          Quality: 40.00   Length: 9
          Ratio: 4.444   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 88.889

alignment_block:
US-08-653-294-14 x AA989542/rev   ..

Align seg 1/1 to reverse of: AA989542 from: 1 to: 402

      2 GluAspArgIleLeuLeuArgTyr 10
        |||::||||||||||||||||||
    353 GAAGAACTCCGGATCTTGCTGCGCTAT 327

seq_name: gb_gss13:AQ440876

seq_documentation_block:
LOCUS   AQ440876   501 bp   DNA   GSS   31-MAR-1999
DEFINITION HS_5098_B2_B04_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=674 Col=8 Row=D, genomic survey sequence.
ACCESSION   AQ440876
VERSION   AQ440876.1 GI:4552215
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 501)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith.K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:

```
                                    http://www.htsc.washington.edu
                                    Plate: 674  row: D column: 8
                                    Seq primer: T7
                                    Class: BAC ends
                                    High quality sequence stop: 501.
FEATURES
        source          1..501
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate-674 Col=8 Row=D"
                        /clone_lib="RPCI-11 Human Male BAC Library"
                        /sex="male"
BASE COUNT     168 a   102 c    80 g   144 t         /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
ORIGIN                                                        7 others

alignment_scores:
        Quality:       40.00                   Length:       10
        Ratio:          4.000                   Gaps:          0
Percent Similarity: 100.000    Percent Identity: 60.000

alignment_block:
US-08-653-294-14 x AQ440876/rev     ..

Align seg 1/1  to reverse of: AQ440876  from: 1  to: 501

    1 ArgGluAspLeuArgIleLeuArgTyr 10
      |||||||||||||||||:::::::||:::
  501 CGTGAAGACCTAAGAGTGAATAATCAGGTTT 472

seq_name: gb_est21:AA975627

seq_documentation_block:
LOCUS      AA975627     505 bp     mRNA         EST      22-MAY-1998
DEFINITION og63b05.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1590993 3'
           similar to gb:L05093 60S RIBOSOMAL PROTEIN L18A (HUMAN);, mRNA
           sequence.
ACCESSION  AA975627
VERSION    AA975627.1   GI:3151419
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 505)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On May 8, 1995 this sequence version replaced gi:801263.
           Contact: Robert Strausberg, Ph.D.
           Tel: (301) 496-1550
           Email: Robert_Strausberg@nih.gov
           Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
           Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: Stratagene, Inc.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html

           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 255.
FEATURES
        source          1..505
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1590993"
                        /clone_lib="NCI_CGAP_Kid6"
                        /sex="mixed"
                        /tissue_type="kidney tumor"
```

```
                                    /lab_host="SOLR (kanamycin resistant)"
                                    /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                                    EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                                    Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
                                    GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                                    CTCGAGTTTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."
BASE COUNT      93 a   115 c   177 g   120 t
ORIGIN

alignment_scores:
        Quality:       40.00                   Length:        9
        Ratio:          4.444                   Gaps:          0
Percent Similarity: 100.000    Percent Identity: 88.889

alignment_block:
US-08-653-294-14 x AA975627/rev     ..

Align seg 1/1  to reverse of: AA975627  from: 1  to: 505

    2 GluAspLeuArgIleLeuArgTyr 10
      |||::|||||||||||||||||||||
  314 GAAGAACTTCGGATTCTGCTGCGCCTAT 288

seq_name: gb_est8:C03945

seq_documentation_block:
LOCUS      C03945        232 bp     mRNA         EST      30-JUL-1996
DEFINITION C03945 Human heart cDNA (YNakamura) Homo sapiens cDNA clone
           3NHC2454, mRNA sequence.
ACCESSION  C03945
VERSION    C03945.1   GI:1467196
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 232)
AUTHORS    Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and
           Nakamura,Y.
TITLE      Construction of a normalized directionally cloned cDNA library from
           adult heart and analysis of 3040 clones by partial sequencing
JOURNAL    Genomics 35 (1), 231-235 (1996)
MEDLINE    96299762
COMMENT    On Oct 24, 1995 this sequence version replaced gi:1040105.
           Contact: Yusuke Nakamura
           Institute of Medical Science
           University of Tokyo
           4-6-1, Shirokaneda, Minato-ku, Tokyo 108,  Japan
           Tel: 81-3-5449-5372
           Fax: 81-3-5449-5433
           Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES            Location/Qualifiers
        source          1..232
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="3NHC2454"
                        /clone_lib="Human heart cDNA (YNakamura)"
                        /dev_stage="adult"
                        /note="Organ: heart; normalized directionally cloned cDNA
                        from adult heart"
BASE COUNT      55 a    77 c    68 g    32 t
ORIGIN

alignment_scores:
        Quality:       39.00                   Length:       10
        Ratio:          4.333                   Gaps:          0
Percent Similarity: 90.000    Percent Identity: 80.000

alignment_block:
US-08-653-294-14 x C03945     ..
```

Align seg 1/1 to: C03945 from: 1 to: 232

```
                 1 ArgGluAspLeuArgIleLeuArgTyr 10
                   ||||||::||||||||||  ||||||||||
                40 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 69
```

seq_name: gb_est10:AA151891

seq_documentation_block:
LOCUS   AA151891   255 bp   mRNA   EST   10-DEC-1996
DEFINITION zo01f06.rl Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:566435 5' similar to gb:M15497_cds1 HLA CLASS I
            HISTOCOMPATIBILITY ANTIGEN, A-24(A-9) A*2401 (HUMAN);, mRNA
            sequence.
ACCESSION   AA151891
VERSION   AA151891.1 GI:1720754
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 255)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
COMMENT   On May 8, 1995 this sequence version replaced gi:800234.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..255
                     /organism="Homo sapiens"
                     /db_xref="GDB:4590888"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:566435"
                     /clone_lib="Stratagene colon (#937204)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. T-84 colonic epithelial cell line. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT   57 a   70 c   75 g   44 t   9 others
ORIGIN
```

alignment_scores:
```
            Quality: 39.00   Length: 10
            Ratio: 4.333   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 80.000
```

alignment_block:
US-08-653-294-14 x AA151891      ..

Align seg 1/1  to: AA151891  from: 1  to: 255

```
                 1 ArgGluAspLeuArgIleLeuArgTyr 10
```

```
                   ||||||::||||||||||  ||||||||||
                77 CGAGAGAACCTGCGGATCGGCNCTCCGCTAC 106
```

seq_name: gb_est21:AA952680

seq_documentation_block:
LOCUS   AA952680   269 bp   mRNA   EST   29-OCT-1998
DEFINITION TENS1864 T. cruzi epimastigote normalized cDNA Library Trypanosoma
            cruzi cDNA clone 1864 5', mRNA sequence.
ACCESSION   AA952680
VERSION   AA952680.1 GI:3115776
KEYWORDS   EST.
SOURCE   Trypanosoma cruzi.
ORGANISM   Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1 (bases 1 to 269)
AUTHORS   Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E.,
            Frasch,A.C.C. and Sanchez,D.O.
TITLE   Gene discovery through expressed sequence tag sequencing in
            trypanosoma cruzi
JOURNAL   Infect. Immun. 66 (11), 5393-5398 (1998)
MEDLINE   99003155
COMMENT   On Jan 17, 1998 this sequence version replaced gi:1900451.
            Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            cp(1650) San Martin, Prov. de BS AS. Argentina
            Tel: (54-1)752-9639 or (54-1)752-0021
            Fax: (54-1)752-0021 or (54-1)752-9639
            Email: dsanchez@inti.gov.ar
            Seq primer: T7.
FEATURES   Location/Qualifiers
     source   1..269
                     /organism="Trypanosoma cruzi"
                     /strain="C1-Brenner"
                     /db_xref="taxon:5693"
                     /clone="1864"
                     /clone_lib="T. cruzi epimastigote normalized cDNA Library"
                     /cell_type="epimastigote"
                     /note="cDNA library constructed with oligo dT primed
                     epimastigote mRNA and cloned in pt7t318D phagemid with
                     modified polylinker (PHARMACIA)"
BASE COUNT   47 a   60 c   89 g   69 t   4 others
ORIGIN
```

alignment_scores:
```
            Quality:      39.00              Length:      10
            Ratio:      3.900               Gaps:        0
Percent Similarity:   100.000       Percent Identity:   60.000
```

alignment_block:
US-08-653-294-14 x AA952680   ..

Align seg 1/1 to: AA952680 from: 1 to: 269

```
                  1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
                    ||||||::||||||::|||::||:|||||||
                177 CGGGAAGAGTTGAGGCTTATTATCCGCTAC 206
```

seq_name: gb_est11:AA263158

seq_documentation_block:
LOCUS   AA263158   283 bp   mRNA   EST   02-JUL-1998
DEFINITION PMY0534 KG1-a Lambda Zap Express cDNA library Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA263158
VERSION   AA263158.1 GI:1898964
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens

```
REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
           Claudio.J.O., Liew.C.C., Dempsey,A.A., Cukerman,E., Stewart,A.K.,
           Na.,E., Atkins,H.I., Iscove,N.N. and Hawley.R.G.
TITLE      Identification of sequence-tagged transcripts differentially
           expressed within the human hematopoietic hierarchy
JOURNAL    Genomics 50 (1), 44-52 (1998)
MEDLINE    98292493
COMMENT    On May 5, 1995 this sequence version replaced gi:797810.
           Contact: Hawley RG
           Oncology Research Laboratories
           The Toronto Hospital
           CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
           Tel: 416 3403834
           Fax: 416 3403453
           Email: r.hawley@utoronto.ca
           Similar to M58636 MHC class I HLA-Bw gene. Clone was randomly
           picked from KG1a primary library.
           Seq primer: 5' GAATTAACCCTCACTAAAGGG 3'
           High quality sequence stop: 283.

FEATURES   Location/Qualifiers
     source    1..283
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="KG1-a Lambda Zap Express cDNA library"
               /cell_type="promyeloblast"
               /cell_line="KG1-a"
               /note="Vector: Lambda Zap Express (Stratagene); Site_1:
               EcoRI; Site_2: XhoI; Unidirectional cloning sites:
               EcoRI-XhoI. mRNA was purified from KG1-a cell line, cDNA
               was synthesized using an XhoI-oligodT linker primer. EcoRI
               adaptors were ligated, followed by digestion with xhoI for
               directional cloning into predigested Lambda Zap Express"

BASE COUNT    64 a    91 c    88 g    40 t
ORIGIN

alignment_scores:
        Quality:    39.00               Length:    10
        Ratio:     4.333                Gaps:       0
Percent Similarity: 90.000    Percent Identity: 80.000

alignment_block:
US-08-653-294-14 x AA263158    ..

Align seg 1/1  to: AA263158  from: 1  to: 283

        1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
          |||||::|||||||||| |||||||||
      120 CGAGAGAACCTGCGGATCGCGCTCCGCTAC 149

seq_name: gb_est6:D82221

seq_documentation_block:
LOCUS      D82221         375 bp    mRNA        EST       09-FEB-1996
DEFINITION HUMHBC4626 Human pancreatic islet Homo sapiens cDNA similar to
           HLA-B, mRNA sequence.
ACCESSION  D82221
VERSION    D82221.1  GI:1183739
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS    Takeda,J.
TITLE      Human pancreatic islet ESTs
JOURNAL    Unpublished (1995)
COMMENT    On Apr 14, 1993 this sequence version replaced gi:785255.
           Contact: Jun Takeda
           Institute for Molecular and Cellular Regulation, Gunma University
           3-39-15 Showa-machi, Maebashi Gunma 371, Japan
```

```
           Tel: 272-20-8856
           Fax: 272-20-8896
           Email: jtakeda@sb.gunma-u.ac.jp.

FEATURES   Location/Qualifiers
     source    1..375
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="Human pancreatic islet"
               /note="Vector: Lambda ZAPII; Site_1: Eco RI: Site_2: Xho
               I; mRNA was prepared from normal adult human islets. cDNA
               was directionally synthesized from the Xho I in the vector
               to the EcoRI site. cDNA was size fractionated to remove
               sequences <1000 bp in size."

BASE COUNT    75 a   124 c   118 g    55 t    3 others
ORIGIN

alignment_scores:
        Quality:    39.00               Length:    10
        Ratio:     4.333                Gaps:       0
Percent Similarity: 90.000    Percent Identity: 80.000

alignment_block:
US-08-653-294-14 x D82221    ..

Align seg 1/1  to: D82221  from: 1  to: 375

        1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
          |||||||::||||||||| |||||||||
      306 CGAGAGAACCTGCGGATCGCGCTCCGNTAC 335

seq_name: gb_gss13:AQ444169

seq_documentation_block:
LOCUS      AQ444169       401 bp   DNA        GSS      31-MAR-1999
DEFINITION GSSTc0231 Trypanosoma cruzi random genomic library Trypanosoma
           cruzi genomic clone G10N2 5', genomic survey sequence.
ACCESSION  AQ444169
VERSION    AQ444169.1  GI:4555633
KEYWORDS   GSS.
SOURCE     Trypanosoma cruzi.
ORGANISM   Trypanosoma cruzi
           Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
           Trypanosoma; Schizotrypanum.
REFERENCE
AUTHORS    1 (bases 1 to 401)
           Sanchez,D.O.
TITLE      Trypanosoma cruzi random genomic sequences
JOURNAL    Unpublished (1999)
COMMENT    Contact: Sanchez D.O.
           Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
           San Martin)
           Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
           CP(1650) San Martin, Prov. de BS AS. Argentina
           Tel: (54-11)4752-0021
           Fax: (54-11)4752-9639
           Email: dsanchez@inti.gov.ar
           Seq primer: T7
           Class: shotgun.

FEATURES   Location/Qualifiers
     source    1..401
               /organism="Trypanosoma cruzi"
               /strain="Cl-Brener"
               /db_xref="taxon:5693"
               /clone="G10N2"
               /clone_lib="Trypanosoma cruzi random genomic library"
               /cell_type="epimastigote"
               /note="Vector: pBS(-) (PHARMACIA)"

BASE COUNT   111 a   129 c    93 g    65 t    3 others
ORIGIN

alignment_scores:
        Quality:    39.00               Length:    10
```

```
                    Ratio: 3.900              Gaps:        0
Percent Similarity: 100.000   Percent Identity: 60.000

alignment_block:
US-08-653-294-14 x AQ44169/rev    ..

Align seg 1/1 to reverse of: AQ44169  from: 1 to: 401

     1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
       ||||||:||||||||:::::::::||||||
   300 CGGGAAGAGTGAGGCTTATTATCCGCTAC 271

seq_name: gb_est37:AI957215

seq_documentation_block:
LOCUS       AI957215        424 bp     mRNA                EST        20-AUG-1999
DEFINITION  u177a10.x1 Sugano mouse kidney mk1a Mus musculus cDNA clone
            IMAGE:2136570 3', mRNA sequence.
ACCESSION   AI957215
VERSION     AI957215.1  GI:5749924
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 424)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187965.
            Other_ESTs: u177a10.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1001246
            Seq primer: custom primer used
            High quality sequence stop: 263.
FEATURES         Location/Qualifiers
     source      1..424
                 /organism="Mus musculus"
                 /strain="C57BL"
                 /db_xref="taxon:10090"
                 /clone_lib="IMAGE:2136570"
                 /clone_lib="Sugano mouse kidney mk1a"
                 /sex="female"
                 /dev_stage="adult"
                 /lab_host="DH10B"
                 /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                 (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                 was primed with an oligo(dT) primer
                 [ATGTGGCCTATTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                 ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested
                 and cloned into distinct DraIII sites of the pME18S-FL3
                 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should
                 be used to isolate the cDNA insert. Size selection was
                 performed to exclude fragments <1.5kb. Library
                 constructed by Dr. Sumio Sugano (University of Tokyo
                 Institute of Medical Science). Custom primers for
                 sequencing; 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                 primer CGACCTGCAGCTCGAGCACA."
BASE COUNT     117 a    116 c    101 g     90 t
ORIGIN
```

```
alignment_scores:
                    Quality:   39.00         Length:      10
                    Ratio:      4.333         Gaps:         0
Percent Similarity: 90.000    Percent Identity: 70.000

alignment_block:
US-08-653-294-14 x AI957215    ..

Align seg 1/1 to: AI957215  from: 1 to: 424

     1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
       |||||||||||:::|||:::|||||||  ||
    88 AGGGAGGACATAAGACTACTACTCAGTTAC 117

seq_name: gb_est10:AA147151

seq_documentation_block:
LOCUS       AA147151        581 bp     mRNA                EST        05-DEC-1996
DEFINITION  zo32d06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:588587 5' similar to gb:M64740 HLA CLASS I HISTOCOMPATIBILITY
            ANTIGEN, A-24(A-9) A*2402 ALPHA (HUMAN):, mRNA sequence.
ACCESSION   AA147151
VERSION     AA147151.1  GI:1716526
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393699.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 272.
FEATURES         Location/Qualifiers
     source      1..581
                 /organism="Homo sapiens"
                 /db_xref="GDB:462889"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:588587"
                 /clone_lib="Stratagene colon (#937204)"
                 /lab_host="SOLR cells (kanamycin resistant)"
                 /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                 Oligo dT; T-84 colonic epithelial cell line. Average
                 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                 sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                 CTCGAGTTTTTTTTTTTTTTTTTT 3 "
BASE COUNT     134 a    162 c    185 g     85 t     15 others
ORIGIN
```

```
alignment_scores:
                    Quality:   39.00         Length:      10
                    Ratio:      4.333         Gaps:         0
Percent Similarity: 90.000    Percent Identity: 80.000
```

alignment_block:
US-08-653-294-14 x AA147151   ..

Align seg 1/1 to: AA147151   from: 1 to: 581

```
        1 ArgGluAspLeuArgIleLeuArgTyr 10
          |||||::||||||||  ||||||||
      152 CGAGAGAACCTGCGGATCNCGCCTCCGCTAC 181
```

seq_name: gb_est26:AI359260

seq_documentation_block:
LOCUS   AI359260   618 bp   mRNA   EST   15-FEB-1999
DEFINITION qy27b07.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2013205 3'
            similar to gb:D32129 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
            AW-66(A-10) A*6601 ALPHA (HUMAN);, mRNA sequence.
ACCESSION   AI359260
VERSION   AI359260.1   GI:4110881
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 618)
AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
JOURNAL   Unpublished (1998)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 691   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES   Location/Qualifiers
     source   1..618
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2013205"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAGTGGAGCGGCCGCAATATCTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT   128 a   171 c   182 g   137 t
ORIGIN
```

alignment_scores:
```
    Quality: 39.00   Length: 10
    Ratio: 4.333   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 80.000
```

alignment_block:
US-08-653-294-14 x AI359260/rev    ..

---

Align seg 1/1  to reverse of: AI359260   from: 1  to: 618

```
        1 ArgGluAspLeuArgIleLeuLeuArgTyr 10
          ||||||:::||||||||  ||||||||
      319 CGAGAGAACCTGCGGATCGCGCCTCCGCTAC 290
```

seq_name: gb_gss13:AQ449604

seq_documentation_block:
LOCUS   AQ449604   715 bp   DNA   GSS   08-APR-1999
DEFINITION 500002D08.x2 CpIOWAM13mp18gDNA1 Cryptosporidium parvum genomic,
            genomic survey sequence.
ACCESSION   AQ449604
VERSION   AQ449604.1   GI:457841
KEYWORDS   GSS.
SOURCE   Cryptosporidium parvum.
ORGANISM   Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE   1 (bases 1 to 715)
AUTHORS   Hyman,R.W., Fung,E., Qin,F., Rowley,D. and Davis,R.W.
TITLE   Cryptosporidium parvum genome sequencing demonstration project
JOURNAL   Unpublished (1999)
COMMENT   On Mar 23, 1999 this sequence version replaced gi:3325323.
            Contact: Hyman, R. W.
            Stanford DNA Sequencing and Technology Center
            Stanford University School of Medicine, Palo Alto
            855 California Avenue, Palo Alto, CA 94304, USA
            Tel: 650 812 1972
            Fax: 650 812 1975
            Email: hyman@sequence.stanford.edu
            For Annotation Data see http://medsfgh.ucsf.edu/id/CpTags/home.html
            Seq primer: M13(-21) forward
            Class: shotgun.
FEATURES   Location/Qualifiers
     source   1..715
                     /organism="Cryptosporidium parvum"
                     /strain="IOWA"
                     /db_xref="taxon:5807"
                     /clone_lib="CpIOWAM13mp18gDNA1"
                     /clone="CpIOWAM13mp18gDNA1"
                     /lab_host="E. coli DH125"
                     /note="Vector: M13mp18; Site_1: Hind III; C. parvum (IOWA
                     isolate) genomic DNA was hydrodynamically sheared to
                     produce fragments having a tight size distribution between
                     1.5 and 3 kb. Adaptors (pGTGACTCA/CAAACCACTGAGTp) were
                     ligated to the randomly sheared gDNA fragments and
                     pAGCTGTTTG linkers were ligated to the Hind III-cleaved
                     M13mp18 vector. The adaptor-containing inserts were
                     annealed and ligated to the vector and transformed into E.
                     coli strain DH125. Recombinant phagemid clones from the
                     first plating of the library were randomly selected for
                     sequence analysis using the M13(-21) forward primer."
BASE COUNT   272 a   92 c   109 g   241 t   1 others
ORIGIN
```

alignment_scores:
```
    Quality:  39.00        Length:       10
    Ratio:     4.333       Gaps:          0
Percent Similarity: 90.000    Percent Identity: 70.000
```

alignment_block:
US-08-653-294-14 x AQ449604   ..

Align seg 1/1 to: AQ449604   from: 1 to: 715

```
        1 ArgGluAspLeuArgIleLeuArgTyr 10
          |||||||||||::::||:||||  |||||
      191 AGAGAAGATCTTAAATTGTGTACCGATAT 220
```

```
           GenCore version 4.5
      Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 8, 2000, 01:29:38 ; Search time 122.56 Seconds
                                         (without alignments)
                                         1.933 Million cell updates/sec

Title:         US-08-653-294-15
Perfect score: 49
Sequence:      1 YRLLIRLDER 10

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      188963 seqs, 23688106 residues

Total number of hits satisfying chosen parameters:    188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :     A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	10	1	W47272	Immunomodulatory p
2	44	89.8	10	1	W47268	Immunomodulatory p
3	44	89.8	10	1	W47270	Immunomodulatory p
4	44	89.8	10	1	W47266	Immunomodulatory p
5	39	79.6	20	1	R92909	HLA-B2702 CTL modu
6	39	79.6	20	1	R92911	HLA-B2702 CTL modu
7	39	79.6	20	1	R92907	HLA-B2702 CTL modu
8	39	79.6	20	1	R95428	HLA-B2702 84-75-84
9	39	79.6	20	1	W33778	Immunomodulating d
10	39	79.6	20	1	W33779	Immunomodulating d
11	39	79.6	20	1	W33792	Peptide B2702.84-7
12	34	69.4	20	1	R92910	HLA-B2702 CTL modu
13	34	69.4	20	1	R92908	HLA-B2702 CTL modu
14	34	69.4	20	1	R95430	HLA-B2702 84-75T/7
15	34	69.4	20	1	W33791	Peptide B2702.84-7
16	34	69.4	20	1	W33793	Peptide B2702.84-7
17	34	69.4	60	1	W93813	Rice anthranilate
18	33	67.3	104	1	W93811	Rice anthranilate
19	33	67.3	577	1	W93815	Rice ASA first iso
20	33	67.3	577	1	W93810	Rice anthranilate
21	32	65.3	451	1	R27642	Human calcium chan
22	30	61.2	319	1	W12377	Regulatory factor
23	30	61.2	381	1	W98786	H. pylori GHPO 121
24	30	61.2	387	1	W89446	A partial gidAl pr
25	30	61.2	621	1	W62842	Helicobacter pylor
26	30	61.2	625	1	W89445	A gidAl protein se
27	30	61.2	775	1	W79193	Human Hrs-2 partia
28	30	61.2	851	1	R41333	113 kD ISGF-3alpha
29	30	61.2	851	1	R72077	Recognition factor
30	30	61.2	851	1	W03166	Human STAT2. New S
31	30	61.2	924	1	W79192	Rat Hrs-2 polypept
32	30	61.2	3079	1	R59926	GAP protein Ira2.
33	30	61.2	3224	1	W54235	Human Nup358 prote
34	29	59.2	6	1	W47264	Immunomodulatory p
35	29	59.2	6	1	W33783	Peptide #4 used in
36	29	59.2	161	1	W98580	H. pylori GHPO 54
37	29	59.2	239	1	W74405	S. aureus gidB pro
38	29	59.2	239	1	W74406	S. aureus gidB pro
39	29	59.2	383	1	W41592	Rat FRAG1 protein.
40	29	59.2	389	1	R15428	3-acylation enzyme
41	29	59.2	485	1	R20796	EHV-4 gC. Nucleic
42	29	59.2	491	1	Y05397	Human TIE ligand N
43	29	59.2	495	1	W89551	Mouse Smad6 protei
44	29	59.2	496	1	W96815	Smad7 protein used
45	29	59.2	568	1	W67442	Bacillus species a

ALIGNMENTS

```
RESULT    1
W47272
ID   W47272 standard; peptide; 10 AA.
AC   W47272;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulatory; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key         Location/Qualifiers
FT   Misc_difference 1..10
FT               /note= "at least one of the amino acids is the
FT                       D-isomer
PN   WO9744052-A1.
PD   27-NOV-1997.
PF   23-APR-1997; U06705.
PR   22-MAY-1996; US-651650.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 98-018220/02.
PT   Novel immunomodulatory peptide-type compound - useful for inhibiting
PT   transplant rejection
PS   Claim 10; Page 36; 41pp; English.
CC   The present sequence is an immunomodulatory peptide, which
CC   comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC   in a pharmaceutical composition together with a subtherapeutic dose
CC   of an immunosuppressant, to extend the period of acceptance of a
CC   transplant from a major histocompatibility complex (MHC) unmatched
CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence  10 AA;

Query Match             100.0%;  Score 49;  DB 1; Length 10;
Best Local Similarity   100.0%;  Pred. No. 0.00025;
Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 YRLLIRLDER 10
          ||||||||||
Db      1 YRLLIRLDER 10

RESULT    2
W47268
ID   W47268 standard; peptide; 10 AA.
AC   W47268;
DT   22-MAY-1998 (first entry)
DE   Immunomodulatory peptide.
KW   Immunomodulatory; Class I HLA-B alpha-1 domain; inhibition;
KW   transplant rejection; treatment; autoimmune disease.
OS   Homo sapiens.
OS   Synthetic.
FH   Key         Location/Qualifiers
FT   Misc_difference 1..10
```

FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI: 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10: Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched
CC donor, i.e. to inhibit transplant rejection. It can also be used in
CC the treatment of autoimmune diseases.
CC Peptides using the D-form amino acids are more effective
CC immunomodulators than their diastereomers or enantiomers.
SQ Sequence 10 AA;

Query Match   89.8%; Score 44; DB 1; Length 10;
Best Local Similarity   90.0%; Pred. No. 0.0027;
Matches   9; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy 1 YRLLIRLDER 10
       ||||||:||
Db 1 YRLLIRLNER 10

RESULT 3
W47270
ID W47270 standard; peptide; 10 AA.
AC W47270;
DT 22-MAY-1998 (first entry)
DE Immunomodulatory peptide.
KW Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW transplant rejection; treatment; autoimmune disease.
OS Homo sapiens.
OS Synthetic.
FH Key   Location/Qualifiers
FT Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI: 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10: Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched
CC donor, i.e. to inhibit transplant rejection. It can also be used in
CC the treatment of autoimmune diseases.
CC Peptides using the D-form amino acids are more effective
CC immunomodulators than their diastereomers or enantiomers.
SQ Sequence 10 AA;

Query Match   89.8%; Score 44; DB 1; Length 10;
Best Local Similarity   90.0%; Pred. No. 0.0027;
Matches   9; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy 1 YRLLIRLDER 10
       ||| |||||||
Db 1 YRLAIRLDER 10

RESULT .4
W47266
ID W47266 standard; peptide; 10 AA.
AC W47266;
DT 22-MAY-1998 (first entry)
DE Immunomodulatory peptide.
KW Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW transplant rejection; treatment; autoimmune disease.
OS Homo sapiens.
OS Synthetic.
FH Key   Location/Qualifiers
FT Misc_difference 1..10
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI: 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10: Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched
CC donor, i.e. to inhibit transplant rejection. It can also be used in
CC the treatment of autoimmune diseases.
CC Peptides using the D-form amino acids are more effective
CC immunomodulators than their diastereomers or enantiomers.
SQ Sequence 10 AA;

Query Match   79.6%; Score 39; DB 1; Length 10;
Best Local Similarity   80.0%; Pred. No. 0.029;
Matches   8; Conservative   1; Mismatches   1; Indels   1; Gaps   0;

Qy 1 YRLLIRLDER 10
       ||| |||:||
Db 1 YRLAIRLNER 10

RESULT 5
R92909
ID R92909 standard; peptide; 20 AA.
AC R92909;
DT 16-MAY-1996 (first entry)
DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppress; graft versus host disorder; transplantation; therapy;
KW class I MHC; HLA-B2702.
OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI: 95-358582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence

```
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match    79.6%;   Score 39;   DB 1;   Length 20;
Best Local Similarity   80.0%;   Pred. No. 0.062;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLDER 10
        ||| |||:||
Db    1 YRLAIRLNER 10

RESULT   6
R92911
ID   R92911 standard; peptide; 20 AA.
AC   R92911;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/84-75).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995;  U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is a dimer of residues 75-84 of the alpha-1 domain of the class I MHC
CC   HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match    79.6%;   Score 39;   DB 1;   Length 20;
Best Local Similarity   80.0%;   Pred. No. 0.062;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLDER 10
        ||| |||:||
Db    1 YRLAIRLNER 10

RESULT   7
R92907
ID   R92907 standard; peptide; 20 AA.
AC   R92907;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995;  U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match    79.6%;   Score 39;   DB 1;   Length 20;
Best Local Similarity   80.0%;   Pred. No. 0.062;
Matches   8;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 YRLLIRLDER 10
        ||| |||:||
Db    1 YRLAIRLNER 10

RESULT   8
R95428
ID   R95428 standard; peptide; 20 AA.
AC   R95428;
DT   12-NOV-1996  (first entry)
DE   HLA-B2702 84-75-84 palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; Cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994;  U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-75-84 palindrome. These sequences can be used to isolate
CC   the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC   protein associated with T-cell activation in mammalian T-cells, and is
CC   also immunologically cross reactive with the heat shock protein Hsc70.
CC   p74 is found in a limited number of cell types, but is particularly
CC   expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC   cell with an amphoteric detergent, and then passed through an affinity
CC   column containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
```

```
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence 20 AA;

Query Match          79.6%;  Score 39;  DB 1;  Length 20;
Best Local Similarity 80.0%;  Pred. No. 0.062;
Matches  8;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy   1 YRLLIRLDER 10
       ||| |||:||
Db   1 YRLAIRLNER 10

RESULT  9
W33778
ID   W33778 standard; peptide; 20 AA.
AC   W33778;
DT   19-JUN-1998 (first entry)
DE   Immunomodulating dimer peptide #1.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 16; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is
CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B -(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or
CC   V; aa77 - D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84= a
CC   hydrophobic or small amino acid; aa82 = R or L; aa83 - G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match          79.6%;  Score 39;  DB 1;  Length 20;
Best Local Similarity 80.0%;  Pred. No. 0.062;
Matches  8;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy   1 YRLLIRLDER 10
       ||| |||:||
Db   1 YRLAIRLNER 10

RESULT  10
W33779
ID   W33779 standard; peptide; 20 AA.
AC   W33779;
```

```
DT   19-JUN-1998 (first entry)
DE   Immunomodulating dimer peptide #2.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain:
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 16; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is
CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B -(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or
CC   V; aa77 - D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84= a
CC   hydrophobic or small amino acid; aa82 = R or L; aa83 - G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alphal domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match          79.6%;  Score 39;  DB 1;  Length 20;
Best Local Similarity 80.0%;  Pred. No. 0.062;
Matches  8;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy   1 YRLLIRLDER 10
       ||| |||:||
Db   1 YRLAIRLNER 10

RESULT  11
W33792
ID   W33792 standard; peptide; 20 AA.
AC   W33792;
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
```

CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula: A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84 = G or R; and aa represents amino
CC   small amino acid: aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match   79.6%; Score 39; DB 1; Length 20;
Best Local Similarity   80.0%; Pred. No. 0.062;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLLIRLDER 10
                ||| ||:||
Db   1 YRLAIRLNER 10

RESULT 12
R92910
ID   R92910 standard; peptide; 20 AA.
AC   R92910;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   69.4%; Score 34; DB 1; Length 20;
Best Local Similarity   70.0%; Pred. No. 0.67;
Matches   7; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLDER 10
                ||| ||:||
Db   1 YRLATRLNER 10

RESULT 13
R92908
ID   R92908 standard; peptide; 20 AA.
AC   R92908;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   69.4%; Score 34; Pred. No. 0.67; Length 20;
Best Local Similarity   70.0%; Pred. No. 0.67;
Matches   7; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   1 YRLLIRLDER 10
                ||| ||:||
Db   1 YRLATRLNER 10

RESULT 14
R95430
ID   R95430 standard; peptide; 20 AA.
AC   R95430;
DT   12-NOV-1996 (first entry)
DE   HLA-B2702 84-75T/75-84T palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-75T/75-84T palindrome. These sequences can be used to
CC   isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface
CC   membrane protein associated with T-cell activation in mammalian T-cells,
CC   and is also immunologically cross reactive with the heat shock protein
CC   Hsc70. p74 is found in a limited number of cell types, but is
CC   particularly expressed on B and T cells. p74 can be isolated by lysis of

CC   a suitable cell with an amphoteric detergent, and then passed through an
CC   affinity column containing a covalently bound HLA-B2702 palindromic
CC   peptide. Compositions comprising the extracellular fragment of p74
CC   combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and
CC   inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.
CC   Candidate compounds can be screened for their effect on the cyolytic
CC   activity of T-cells, by combining them with the extracellular portion of
CC   p74 and determining the amount of binding between the candidate compound
CC   and p74. Modulation of CTL activity can be inhibited in a cellular
CC   composition containing T-cells and antigen presenting cells (APCs), by
CC   adding to the mix the extracellular portion of p74, in an amount
CC   sufficient to compete with p74 for the binding of the p74 ligand.
SQ   Sequence   20 AA;

Query Match   69.4%; Score 34; DB 1; Length 20;
Best Local Similarity   77.8%; Pred. No. 0.67;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLLIRLDE 9
            ||| ||| :||
Db   1 YRLAIRLNE 9

RESULT 15
W33791
ID   W33791 standard; peptide; 20 AA.
AC   W33791;
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75T/75-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B -
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa77 - E or V; aa77 -
CC   D. S or N: aa79 - R or G: aa80 - I or N: aa81, aa83 - G or R; and aa
CC   small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   20 AA;

Query Match   69.4%; Score 34; DB 1; Length 20;
Best Local Similarity   70.0%; Pred. No. 0.67;
Matches   7; Conservative   2; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLLIRLDER 10

Db   1 YRLATRLNER 10
            ||| ||:||

Search completed: February 8, 2000, 01:29:38
Job time: 1750 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   February 7, 2000, 11:54:24 ; Search time 117.7 Seconds
                                        (without alignments)
                                        4.008 Million cell updates/sec

Title:            US-08-653-294-13
Perfect score:    49
Sequence:         1 YRLAIRLDER 10

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Database :        PIR_62:*
                  1: pir1:*
                  2: pir2:*
                  3: pir3:*
                  4: pir4:*

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters:    142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	69.4	1154	2	S43275	hypothetical prote
2	34	69.4	1154	2	S43277	hypothetical prote
3	33	67.3	321	2	C64941	hypothetical prote
4	33	67.3	554	1	F70548	probable menD prot
5	32	65.3	151	2	C71113	probable frxA prot
6	32	65.3	185	2	S74416	hypothetical prote
7	32	65.3	318	2	A55429	11-cis retinol deh
8	32	65.3	319	2	I45845	11-cis-retinol deh
9	32	65.3	349	1	RGECGL	nitrogen regulatio
10	32	65.3	349	2	A24114	nitrogen regulatio
11	32	65.3	464	2	D42902	probable acetate-
12	32	65.3	464	2	S59898	kynureninase (EC 3
13	32	65.3	546	2	S61294	heat-shock protein
14	32	65.3	905	2	H71731	DNA gyrase chain A
15	31	63.3	236	2	B72299	sugar-phosphate al
16	31	63.3	333	2	T05643	hypothetical prote
17	31	63.3	463	2	B72500	probable seryl-tRN
18	31	63.3	485	1	B45343	glycoprotein gp13
19	31	63.3	510	2	A70580	probable UDP-N-ace
20	31	63.3	544	2	B34089	carboxylesterase (
21	31	63.3	654	2	S58820	DNA-directed RNA p
22	31	63.3	705	2	S54521	probable membrane
23	31	63.3	741	2	T16992	ethylene receptor
24	31	63.3	1063	2	A33830	cation efflux syst
25	31	63.3	1063	2	JC4700	cadmium, zinc, cob
26	31	63.3	1214	2	JC2069	zinc-finger protei
27	31	63.3	1524	2	S68553	surface layer prot
28	31	63.3	1699	2	T14074	complement compone
29	31	63.3	2067	2	A42854	probable spindle p
30	30	61.2	124	2	E70008	conserved hypothet
31	30	61.2	125	2	D72544	hypothetical prote
32	30	61.2	141	1	F70457	hypothetical prote
33	30	61.2	204	2	C40899	hypothetical prote
34	30	61.2	252	2	H64752	probable transcrip
35	30	61.2	303	2	H71277	probable DNA adeni
36	30	61.2	312	2	T15371	hypothetical prote
37	30	61.2	336	2	H70693	hypothetical prote
38	30	61.2	465	2	JE0369	histidine acid pho
39	30	61.2	506	1	S37583	RING finger protei
40	30	61.2	513	1	TVHURF	ret finger protein
41	30	61.2	540	2	S76869	hypothetical prote
42	30	61.2	545	2	T00485	probable phosphor1
43	30	61.2	547	2	A56575	puff-specific nucl
44	30	61.2	610	2	G69130	conserved hypothet
45	30	61.2	723	2	T14605	probable cell divi

ALIGNMENTS

```
RESULT 1
S43275
hypothetical protein 2 - Neurospora crassa retrotransposon Tad1-1
C;Species: Neurospora crassa
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Sep-1997
C;Accession: S43275
R;Cambareri, E.B.; Helber, J.; Kinsey, J.A.
Mol. Gen. Genet. 242, 658-665, 1994
A;Title: Tad1-1, an active LINE-like element of Neurospora crassa.
A;Reference number: S43274; MUID:94203179
A;Accession: S43275
A;Molecule type: DNA
A;Residues: 1-1154 <CAM>
A;Cross-references: EMBL:L25662; NID:g409759; PID:g409761
C;Genetics:
A;Mobile element: retrotransposon Tad1-1

Query Match             69.4%;   Score 34;   DB 2;   Length 1154;
Best Local Similarity   66.7%;          Pred. No. 51;
Matches    6;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy      1 YRLAIRLDE 9
          |||:|:|:|
Db   1136 YRLAVELEE 1144

RESULT 2
S43277
hypothetical protein 2 - Neurospora crassa retrotransposon Tad3-2
C;Species: Neurospora crassa
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Sep-1997
C;Accession: S43277
R;Cambareri, E.B.; Helber, J.; Kinsey, J.A.
Mol. Gen. Genet. 242, 658-665, 1994
A;Title: Tad1-1, an active LINE-like element of Neurospora crassa.
A;Reference number: S43277; MUID:94203179
A;Accession: S43277
A;Molecule type: DNA
A;Residues: 1-1154 <CAM>
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Cross-references: EMBL:L25663; NID:g409762; PID:g409764
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993

Query Match             69.4%;   Score 34;   DB 2;   Length 1154;
Best Local Similarity   66.7%;          Pred. No. 51;
Matches    6;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy      1 YRLAIRLDE 9
          |||:|:|:|
Db   1136 YRLAVELEE 1144
```

RESULT 3
C64941
hypothetical protein b1803 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: C64941
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
Science 277, 1453-1462, 1997
A:Reference number: A64720; MUID:97426617
A:Accession: C64941
A:Status: preliminary: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-321 <BLAT>
A:Cross-references: GB:AE000274; GB:U00096; NID:g1788089; PIDN:AAC74873.1; PID:g1788104;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred
F:11-225/Domain: cytochrome-b5 reductase homology <CBR>
F:254-309/Domain: ferredoxin [2Fe-2S] homology <FER>

```
Query Match              67.3%;   Score 33;   DB 2;   Length 321;
Best Local Similarity    55.6%;   Pred. No. 22;
Matches     5;   Conservative    4;   Mismatches    0;   Indels    0;   Gaps    0;

Qy      1 YRLAIRLDE 9
            |:|:|:|:|
Db     70 YQIAVRLEE 78
```

RESULT 4
F70548
probable menD protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: F70548
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70548
A:Status: preliminary: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-554 <COL>
A:Cross-references: GB:Z95558; GB:AL123456; NID:g3261781; PID:e316800; PID:g2114017
C:Genetics:
A:Gene: menD
C:Superfamily: menD protein

```
Query Match              67.3%;   Score 33;   DB 1;   Length 554;
Best Local Similarity    66.7%;   Pred. No. 39;
Matches     6;   Conservative    2;   Mismatches    2;   Indels    1;   Gaps    0;

Qy      2 RLAIRLDER 10
            || :|:|||
Db     48 RLHVRIDER 56
```

RESULT 5
C71113
probable frxA protein - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: C71113
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A:Reference number: A71000; MUID:98344137
A:Accession: C71113
A:Status: preliminary: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <KAW>
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030708; PID:g3257082
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0674

```
Query Match              65.3%;   Score 32;   DB 2;   Length 151;
Best Local Similarity    77.8%;   Pred. No. 16;
Matches     7;   Conservative    0;   Mismatches    2;   Indels    0;   Gaps    0;

Qy      2 RLAIRLDER 10
            || |||||
Db     96 RLLIEDLER 104
```

RESULT 6
S74416
hypothetical protein sll0687 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S74416
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, Y.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S74416
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-185 <KAN>
A:Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PID:d1010985; PID:g100119
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

```
Query Match              65.3%;   Score 32;   DB 2;   Length 185;
Best Local Similarity    60.0%;   Pred. No. 20;
Matches     6;   Conservative    3;   Mismatches    1;   Indels    0;   Gaps    0;

Qy      1 YRLAIRLDE 10
            |||:|: :|
Db     35 YRLALRILQR 44
```

RESULT 7
A55429
11-cis retinol dehydrogenase (EC 1.1.1.-) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 29-Sep-1999
C:Accession: A55429
R:Simon, A.; Hellman, U.; Wernstedt, C.; Eriksson, U.
J. Biol. Chem. 270, 1107-1112, 1995
A:Title: The retinal pigment epithelial-specific 11-cis retinol dehydrogenase belongs
A:Reference number: A55429; MUID:95138097
A:Accession: A55429
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-318 <SIM>
A:Cross-references: GB:X82262; NID:g663170; PIDN:CAA57715.1; PID:g663171
C:Superfamily: retinol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: membrane protein; NAD; oxidoreductase
F:29-206/Domain: short-chain alcohol dehydrogenase homology <SADH>

```
Query Match            65.3%;   Score 32;  DB 2;   Length 318;
Best Local Similarity  75.0%;   Pred. No. 35;
Matches   6;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    3 LAIRLDER 10
          ||:|||:|
Db   44 LALRLDQR 51
```

RESULT 8
I45845
11-cis-retinol dehydrogenase (EC 1.1.1.-) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 29-Sep-1999
C:Accession: I45845
R:Driessen, C.A.; Janssen, B.P.; Winkens, H.J.; van Vugt, A.H.; de Leeuw, T.L.; Janssen,
Invest. Ophthalmol. Vis. Sci. 36, 1988-1996, 1995
A:Title: Cloning and expression of a cDNA encoding bovine retinal pigment epithelial 11-
A:Reference number: I45845
A:Accession: I45845
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-319 <DRI>
A:Cross-references: GB:L36533; NID:g1054530; PIDN:AAA80694.1; PID:g1054531
C:Superfamily: retinol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase
F:30-207/Domain: short-chain alcohol dehydrogenase homology <SADH>

```
Query Match            65.3%;   Score 32;  DB 2;   Length 319;
Best Local Similarity  75.0%;   Pred. No. 35;
Matches   6;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    3 LAIRLDER 10
          ||:|||:|
Db   45 LALRLDQR 52
```

RESULT 9
RGECGL
nitrogen regulation protein II (EC 2.7.3.-) ntrB - Escherichia coli
N:Alternate names: regulatory protein glnL
C:Species: Escherichia coli
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C:Accession: A30377; S40814; B23970; H65191; A39765; Q00553
R:Miranda-Rios, J.; Sanchez-Pescador, R.; Urdea, M.; Covarrubias, A.A.
Nucleic Acids Res. 15, 2757-2770, 1987
A:Title: The complete nucleotide sequence of the glnALG operon of Escherichia coli K12.
A:Reference number: A30377; MUID:87174797
A:Accession: A30377
A:Molecule type: DNA
A:Residues: 1-349 <MIR>
A:Cross-references: EMBL:X05173; NID:g41562; PIDN:CAA28807.1; PID:g41564
A:Experimental source: K-12
R:Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 21, 3391-3398, 1993
A:Title: Analysis of the Escherichia coli genome. III. DNA sequence of the region from 8
A:Reference number: S40802; MUID:93347969
A:Accession: S40814
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-349 <PLU>
A:Cross-references: EMBL:L19201; NID:g304961; PIDN:AAB03003.1; PID:g304974
A:Experimental source: strain K-12, substrain MG1655
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1993
R:Rocha, M.; Vazquez, M.; Garciarrubio, A.; Covarrubias, A.A.
Gene 37, 91-99, 1985
A:Title: Nucleotide sequence of the glnA-glnL intercistronic region of Escherichia coli.
A:Reference number: A91533; MUID:86031370
A:Accession: B23970
A:Molecule type: DNA
A:Residues: 1-24 <ROC>
A:Cross-references: GB:K02176; GB:M11581; NID:g146160; PIDN:AAA23881.1; PID:g146162

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H65191
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-349 <BLAT>
A:Cross-references: GB:AE000462; GB:U00096; NID:g1790295; PIDN:AAC76866.1; PID:g17903
A:Experimental source: strain K-12, substrain MG1655
R:Ninfa, A.J.; Bennett, R.L.
J. Biol. Chem. 266, 6888-6893, 1991
A:Title: Identification of the site of autophosphorylation of the bacterial protein k
A:Reference number: A39765; MUID:91201336
A:Accession: A39765
A:Status: preliminary
A:Molecule type: protein
A:Residues: 2-11;136-142;158-162,'X',164-169 <NIN>
C:Genetics:
A:Gene: glnL; ntrB
A:Map position: 87 min
C:Function:
A:Description: de-uridylylated P-II forms a complex with nitrogen regulation protein
the uridylylated form of P-II does not complex with ntrB; free ntrB phosphorylates n1
A:Note: phosphorylated nitrogen regulation protein I (ntrC) activates transcription o
C:Superfamily: glnL regulatory protein II; sensor histidine kinase homology
C:Keywords: ATP; autophosphorylation; phosphohistidine; phosphoprotein; phosphotransf
F:104-346/Domain: sensor histidine kinase homology <SHK>
F:139/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status pred1
F:329/Binding site: ATP (Lys) #status predicted

```
Query Match            65.3%;   Score 32;  DB 1;   Length 349;
Best Local Similarity  75.0%;   Pred. No. 39;
Matches   6;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 YRLAIRLD 8
          ||||:|:|
Db  277 YRLAARID 284
```

RESULT 10
A24114
nitrogen regulation protein II (EC 2.7.3.-) ntrB - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 22-Jul-1987 #sequence_revision 22-Jul-1987 #text_change 20-Aug-1999
C:Accession: A24114
R:MacFarlane, S.A.; Merrick, M.
Nucleic Acids Res. 13, 7591-7607, 1985
A:Reference number: A24114; MUID:86067184
A:Accession: A24114
A:Molecule type: <MAC>
A:Residues: 1-349 <MAC>
A:Cross-references: GB:X03146; NID:g43893; PIDN:CAA26923.1; PID:g43895
C:Genetics:
A:Gene: ntrB
C:Superfamily: glnL regulatory protein II; sensor histidine kinase homology
C:Keywords: ATP; autophosphorylation; phosphohistidine; phosphoprotein; phosphotransf
F:104-346/Domain: sensor histidine kinase homology <SHK>
F:139/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status pred1
F:329/Binding site: ATP (Lys) #status predicted

```
Query Match            65.3%;   Score 32;  DB 2;   Length 349;
Best Local Similarity  75.0%;   Pred. No. 39;
Matches   6;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 YRLAIRLD 8
          ||||:|:|
Db  277 YRLAARID 284
```

RESULT 11
D42902
probable acetate--CoA ligase (EC 6.2.1.1) - Pseudomonas aeruginosa (fragment)
N;Alternate names: probable acetyl-CoA sythetase
C;Species: Pseudomonas aeruginosa
A;Molecule type: DNA
C;Date: 04-Mar-1993 #sequence_revision 24-Oct-1997 #text_change 24-Oct-1997
C;Accession: S27604; D42902
R;Steele, M.I.; Lorenz, D.; Hatter, K.; Parks, A.; Sokatch, J.R.
submitted to the EMBL Data Library, July 1992
A;Description: Characterization of the mmsAB operon of Pseudomonas aeruginosa PAO encodi
A;Reference number: S27601
A;Accession: S27604
A;Molecule type: DNA
A;Residues: 1-464 <STE>
A;Cross-references: EMBL:M84911; NID:g151360; PID:g551933
R;Steele, M.I.; Lorenz, D.; Hatter, K.; Park, A.; Sokatch, J.R.
J. Biol. Chem. 267, 13585-13592, 1992
A;Title: Characterization of the mmsAB operon of Pseudomonas aeruginosa PAO encoding met
A;Reference number: A42902; MUID:92317087
A;Accession: D42902
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-79 <ST2>
A;Experimental source: PAO, ATCC 15692
A;Note: sequence extracted from NCBI backbone (NCBIN:107704, NCBIP:107709)
C;Superfamily: acetate--CoA ligase homology
C;Keywords: acid-thiol ligase
F;105-464/Domain: acetate--CoA ligase homology (fragment) <ACL>

Query Match
Best Local Similarity   65.3%;   Score 32;   DB 2;   Length 464;
Matches   5;   Conservative   3;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 YRLAIRLDE 9
         | ||:||:|:
Db   146 YELALRIDD 154

RESULT 12
S59898
kynureninase (EC 3.7.1.3) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
C;Accession: S59898
R;Takeuchi, F.; Tsubouchi, R.; Yoshino, M.; Shibata, Y.
Biochim. Biophys. Acta 1252, 185-188, 1995
A;Title: Amino-acid sequence of rat liver kynureninase.
A;Reference number: S59898; MUID:96049498
A;Accession: S59898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-464 <TAK>
C;Keywords: hydrolase

Query Match
Best Local Similarity   65.3%;   Score 32;   DB 2;   Length 464;
Matches   6;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   2 RLAIRLDE 9
         |:|:||||
Db   28 RVALRLDE 35

RESULT 13
S61294
heat-shock protein - Pyrococcus sp.
C;Species: Pyrococcus sp.
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Aug-1999
C;Accession: S61294
R;Izawa, Y.; Kakihara, H.; Takagi, M.; Imanaka, T.
submitted to the EMBL Data Library, March 1994

A;Description: Cloning and analysis of the heat shock protein gene from a new hyperth
A;Reference number: S61294
A;Accession: S61294
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <IZA>
A;Cross-references: EMBL:D29672; NID:g473964; PIDN:BAA06143.1; PID:d1006705; PID:g473
C;Superfamily: molecular chaperone t-complex-type
C;Keywords: heat shock

Query Match
Best Local Similarity   100.0%;   Score 32;   DB 2;   Length 546;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   3 LAIRLDE 9
         |||||||
Db   418 LAIRLDE 424

RESULT 14
H71731
DNA gyrase chain A (gyrA) RP206 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 16-Jul-1999
C;Accession: H71731
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: H71731
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-905 <AND>
A;Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14671.1; PID:e134
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: gyrA; RP206
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomeras
F;1-239/Domain: phage T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology <T

Query Match
Best Local Similarity   75.0%;   Score 32;   DB 2;   Pred. No. 1e+02;
Matches   6;   Conservative   2;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2 RLAIRLDE 9
         ::|||||
Db   665 KIAIRLDE 672

RESULT 15
B72299
sugar-phosphate aldolase - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: B72299
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316
A;Accession: B72299
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-236 <ARN>
A;Cross-references: GB:AE001767; GB:AE000512; NID:g4981611; PID:g4981616; TIGR:TM1072
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1072

Qy   2 RLAIRLDE 9
         :|||||
Db   665 KIAIRLDE 672

Query Match   63.3%; Score 31; DB 2; Length 236;
Best Local Similarity 62.5%; Pred. No. 42;
Matches   5; Conservative   3; Mismatches   0; Indels   0; Gaps   0;

Qy   3 LAIRLDER 10
        ::|||||
Db   28 ISVRLDER 35

Search completed: February 7, 2000, 11:54:25
Job time: 24335 sec

THIS PAGE BLANK (USPTO)

```
         GenCore version 4.5
      Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   February 8, 2000, 00:59:52 ; Search time 63.71 Seconds
                                        (without alignments)
                                        4.688 Million cell updates/sec

Title:           US-08-653-294-13
Perfect score:   49
Sequence:        1 YRLAIRLDER 10

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters:     82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :       SwissProt_38:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID		Description
1	33	67.3	321	1	YEAX_ECOLI	P76254	escherichia
2	33	65.3	318	1	RDH1_BOVIN	Q27979	bos taurus
3	32	65.3	349	1	NTRB_ECOLI	P06712	escherichia
4	32	65.3	349	1	NTRB_KLEPN	P06218	klebsiella
5	32	65.3	464	1	KYNU_RAT	P70712	rattus norv
6	32	65.3	464	1	YMMS_PSEAE	P28812	pseudomonas
7	32	65.3	546	1	THS_PYRKO	Q52500	pyrococcus
8	32	65.3	548	1	THSA_PYRKO	024729	thermococcu
9	32	65.3	905	1	GYRA_RICPR	P41080	rickettsia
10	31	63.3	485	1	VGLC_HSVE4	P22596	equine herp
11	31	63.3	510	1	MURF_MYCTU	O06220	mycobacteri
12	31	63.3	544	1	ESTP_DROME	P18167	drosophila
13	31	63.3	654	1	RPC3_YEAST	P32349	saccharomyc
14	31	63.3	705	1	YM37_YEAST	Q03824	saccharomyc
15	31	63.3	1014	1	MMLB_MYCLE	O06079	mycobacteri
16	31	63.3	1063	1	C2CA_ALCEU	P13511	alcaligenes
17	31	63.3	1063	1	C2CA_ALCSP	P94177	alcaligenes
18	31	63.3	1214	1	BR14_HUMAN	P55201	homo sapien
19	31	63.3	2067	1	BIMB_EMENI	P33144	emericella
20	30	61.2	252	1	YAGI_ECOLI	P77300	escherichia
21	30	61.2	303	1	DMA_TREPA	033844	treponema p
22	30	61.2	513	1	RFP_HUMAN	P14373	homo sapien
23	30	61.2	522	1	RFP_MOUSE	Q62158	mus musculu
24	30	61.2	547	1	BX42_DROME	P39736	drosophila
25	30	61.2	663	1	TERM_ADEB3	055439	bovine aden
26	30	61.2	1176	1	NIR_NEUCR	P38681	neurospora
27	30	61.2	3491	1	ERY1_SACER	Q03131	saccharopol
28	30	61.2	3587	1	SRF1_BACSU	P27206	bacillus su
29	29	59.2	108	1	YJT9_YEAST	P39534	saccharomyc
30	29	59.2	158	1	YCBM_BACSU	P42245	bacillus su
31	29	59.2	189	1	VH02_VACCC	P20496	vaccinia vi
32	29	59.2	189	1	VH02_VACCV	P08583	vaccinia vi
33	29	59.2	189	1	VH02_VARV	P33061	variola vir
34	29	59.2	210	1	FLPA_ARCFU	028192	archaeoglob
35	29	59.2	246	1	TRYP_MOUSE	P07146	mus musculu
36	29	59.2	279	1	LEP3_ERWCA	P31712	erwinia car
37	29	59.2	294	1	RL5A_SCHPO	P52822	schizosacch
38	29	59.2	294	1	RL5B_SCHPO	074306	schizosacch
39	29	59.2	297	1	RL5_HELAN	065353	helianthus
40	29	59.2	299	1	RL5_BOMMO	076190	bombyx mori
41	29	59.2	306	1	MK16_YEAST	P10962	saccharomyc
42	29	59.2	334	1	RUVB_THEMA	Q56313	thermotoga
43	29	59.2	377	1	CAH1_CHLRE	P20507	chlamydomon
44	29	59.2	409	1	METK_SYNY3	P72871	synechocyst
45	29	59.2	428	1	YUR1_YEAST	P26725	saccharomyc

ALIGNMENTS

```
RESULT 1
YEAX_ECOLI
ID   YEAX_ECOLI     STANDARD;     PRT;    321 AA.
AC   P76254; O07972; O07970;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PUTATIVE DIOXYGENASE BETA SUBUNIT YEAX (EC 1.-.-.-).
GN   YEAX.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97251358.
RA   ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA   KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,
RA   MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,
RA   NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA   SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,
RA   YAMAMOTO Y., HORIUCHI T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
     -!- COFACTOR: FMN (BY SIMILARITY).
     -!- SUBUNIT: PROBABLE HETERODIMER OF YEAW AND YEAX.
     -!- SIMILARITY: IN THE C-TERMINAL, BELONGS TO THE 2FE2S PLANT-TYPE
CC       FERREDOXIN FAMILY.
     -!- SIMILARITY: BELONGS TO THE PDR/VANB FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE000274; AAC74873.1; -.
DR   EMBL; D90824; CAB21531.1; -.
DR   EMBL; D90823; CAB21524.1; -.
DR   HSSP; P33164; 2PIA.
DR   ECOGENE; EG13510; YEAX.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PFAM; PF00111; fer2; 1.
DR   PFAM; PF00175; oxidored_fad; 1.
```

```
KW   Hypothetical protein; Oxidoreductase; Flavoprotein; FMN; NAD;
KW   Iron-sulfur; Electron transport.
FT   NP_BIND       6    103       FMN (BY SIMILARITY).
FT   NP_BIND     113    226       NAD (BY SIMILARITY).
FT   METAL       270    270       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT   METAL       275    275       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT   METAL       278    278       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT   METAL       309    309       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ   SEQUENCE   321 AA;  35661 MW;  9E85CC68 CRC32;

Query Match             67.3%;  Score 33;  DB 1;  Length 321;
Best Local Similarity   55.6%;  Pred. No. 11;
Matches   5;  Conservative  4;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    1 YRLAIRLDE 9
         |:|:|||:|
Db   70 YQIAVRLEE 78

RESULT 2
RDH1_BOVIN
ID   RDH1_BOVIN     STANDARD;      PRT;    318 AA.
AC   Q27979;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   11-CIS RETINOL DEHYDROGENASE (EC 1.1.1.105) (11-CIS RDH) (P32).
GN   RDH1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE;
RX   MEDLINE; 95138097.
RA   SIMON A., HELLMAN U., WERNSTEDT C., ERIKSSON U.;
RT   "The retinal pigment epithelial-specific 11-cis retinol dehydrogenase
RT   belongs to the family of short chain alcohol dehydrogenases.";
RL   J. Biol. Chem. 270:1107-1112(1995).
-!- FUNCTION: STEREOSPECIFIC 11-CIS RETINOL DEHYDROGENASE, WHICH
    CATALYZES THE FINAL STEP IN THE BIOSYNTHESIS OF 11-CIS
    RETINALDEHYDE, THE UNIVERSAL CHROMOPHORE OF VISUAL PIGMENTS.
    ACTIVE IN THE PRESENCE OF NAD+ AS COFACTOR BUT NOT IN THE PRESENCE
    OF NADP.
-!- CATALYTIC ACTIVITY: RETINOL + NAD(+) = RETINAL + NADH.
-!- PATHWAY: CATALYZES THE PRIMARY AND RATE-LIMITING STEP IN RETINOIC
    ACID SYNTHESIS.
-!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
    FAMILY (SDR).
```

```
DR   EMBL; X82262; CAA57715.1; -.
DR   HSSP; P14061; 1FDW.
DR   PROSITE; PS00061; ADH_SHORT; FALSE_NEG.
DR   PFAM; PF00106; adh_short; 1.
KW   Oxidoreductase; NAD.
FT   NP_BIND      32     56       NADP (BY SIMILARITY).
FT   ACT_SITE    175    175       BY SIMILARITY.
SQ   SEQUENCE   318 AA;  35036 MW;  7022A583 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 318;
Best Local Similarity   75.0%;  Pred. No. 18;
```

```
Matches   6;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    3 LAIRLDER 10
         ||:|||:|
Db   44 LALRLDQR 51

RESULT 3
NTRB_ECOLI
ID   NTRB_ECOLI     STANDARD;      PRT;    349 AA.
AC   P06712;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NITROGEN REGULATION PROTEIN NR(II) (EC 2.7.3.-).
GN   GLNL OR NTRB OR GLNR.
OS   Escherichia coli.
OC   Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 87174797.
RA   MIRANDA-RIOS J., SANCHEZ-PESCADOR R., URDEA M., COVARRUBIAS A.A.;
RT   "The complete nucleotide sequence of the glnALG operon of Escherichia
RT   coli K12.";
RL   Nucleic Acids Res. 15:2757-2770(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 93347969.
RA   PLUNKETT G. III., BURLAND V.D., DANIELS D.L., BLATTNER F.R.;
RT   "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT   region from 87.2 to 89.2 minutes.";
RL   Nucleic Acids Res. 21:3391-3398(1993).
RN   [3]
RP   SEQUENCE OF 1-22 FROM N.A.
RX   MEDLINE; 85006814.
RA   UENO-NISHIO S., MANGO S., REITZER L.J., MAGASANIK B.;
RT   "Identification and regulation of the glnL operator-promoter of the
RT   complex glnALG operon of Escherichia coli.";
RL   J. Bacteriol. 160:379-384(1984).
RN   [4]
RP   SEQUENCE OF 1-24 FROM N.A.
RX   MEDLINE; 86031370.
RA   ROCHA M., VAZQUEZ M., GARCIARRUBIO A., COVARRUBIAS A.A.;
RT   "Nucleotide sequence of the glnA-glnL intercistronic region of
RT   Escherichia coli.";
RL   Gene 37:91-99(1985).
RN   [5]
RP   PHOSPHORYLATION SITE.
RX   MEDLINE; 91201336.
RA   NINFA A.J., BENNETT R.L.;
RT   "Identification of the site of autophosphorylation of the bacterial
RT   protein kinase/phosphatase NRII.";
RL   J. Biol. Chem. 266:6888-6893(1991).
-!- FUNCTION: NTRB ACTS AS A SIGNAL TRANSDUCER WHICH RESPONDS TO THE
    NITROGEN LEVEL OF CELL AND MODULATES THE ACTIVITY OF NTRC. IN
    NITROGEN LIMITATION NTRB ACTIVATES NTRC BY PHOSPHORYLATING IT,
    WHILE IN NITROGEN EXCESS NTRC IS DEPHOSPHORYLATED AND CONSEQUENTLY
    INACTIVATED BY NTRB.
-!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
    KINASES.
```

```
DR   EMBL; X05173; CAA28807.1; -.
```

```
DR   EMBL; K02176; AAA23881.1; -.
DR   EMBL; L19201; AAB03003.1; -.
DR   EMBL; AE000462; AAC76866.1; -.
DR   PIR; Q00553; RGECGL.
DR   PIR; B23970.
DR   PIR; S40814; S40814.
DR   ECOGENE; EG10387; GLNL.
DR   PFAM; PF00512; signal; 1.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Nitrogen fixation; ATP-binding.
FT   DOMAIN      116    349       TRANSMITTER DOMAIN (POTENTIAL).
FT   MOD_RES     139    139       PHOSPHORYLATION (AUTO).
FT   BINDING     329    329       ATP (BY SIMILARITY).
SQ   SEQUENCE   349 AA;  38556 MW;  6A017919 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 349;
Best Local Similarity   75.0%;  Pred. No. 19;
Matches    6;  Conservative   1;  Mismatches   1;  Indels   1;  Gaps   0;

Qy      1 YRLAIRLD 8
          |||| |:|
Db    277 YRLAARID 284

RESULT 4
NTRB_KLEPN
ID   NTRB_KLEPN     STANDARD;      PRT;   349 AA.
AC   P06218;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   NITROGEN REGULATION PROTEIN NTRB (EC 2.7.3.-).
GN   NTRB.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86067184.
RA   MCFARLANE S.A., MERRICK M.J.;
RT   "The nucleotide sequence of the nitrogen regulation gene ntrB and the
RT   glnA-ntrBC intergenic region of Klebsiella pneumoniae.";
RL   Nucleic Acids Res. 13:7591-7606(1985).
CC   -!- FUNCTION: NTRB ACTS AS A SIGNAL TRANSDUCER WHICH RESPONDS TO THE
CC       NITROGEN LEVEL OF CELL AND MODULATES THE ACTIVITY OF NTRC. IN
CC       NITROGEN LIMITATION NTRB ACTIVATES NTRC BY PHOSPHORYLATING IT,
CC       WHILE IN NITROGEN EXCESS NTRC IS DEPHOSPHORYLATED AND CONSEQUENTLY
CC       INACTIVATED BY NTRB.
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC       KINASES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by  non-profit  institutions as  long  as its content is in  no  way
CC   modified and this statement is not removed.  Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X03146; CAA26923.1; -.
DR   PIR; A24114; A24114.
DR   PFAM; PF00512; signal; 1.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Nitrogen fixation; ATP-binding.
FT   DOMAIN      116    349       TRANSMITTER DOMAIN (POTENTIAL).
FT   MOD_RES     139    139       PHOSPHORYLATION (AUTO) (BY SIMILARITY).
FT   BINDING     329    329       ATP (BY SIMILARITY).
SQ   SEQUENCE   349 AA;  38409 MW;  4BAC1813 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 349;
Best Local Similarity   75.0%;  Pred. No. 19;
Matches    6;  Conservative   1;  Mismatches   1;  Indels   1;  Gaps   0;

Qy      1 YRLAIRLD 8
          |||| |:|
Db    277 YRLAARID 284

RESULT 5
KINU_RAT
ID   KINU_RAT       STANDARD;      PRT;   464 AA.
AC   P70712;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   KYNURENINASE (EC 3.7.1.3) (L-KYNURENINE HYDROLASE).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 97324088.
RA   TOMA S., NAKAMURA M., TONE S., OKUNO E., KIDO R., BRETON J.,
RA   AVANZI N., COZZI L., SPECIALE C., MOSTARDINI M., GATTI S., BENATTI L.;
RT   "Cloning and recombinant expression of rat and human kynureninase.";
RL   FEBS Lett. 408:5-10(1997).
CC   -!- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYNURENINE AND L-3-
CC       HYDROXYKYNURENINE INTO ANTHRANILIC AND 3-HYDROXYANTHRANILIC ACIDS,
CC       RESPECTIVELY.
CC   -!- CATALYTIC ACTIVITY: L-KYNURENINE + H(2)O = ANTHRANILATE +
CC       L-ALANINE.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM
CC       TRYPTOPHAN THROUGH THE KYNURENINE PATHWAY.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED
CC       TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by  non-profit  institutions as  long  as  its content is in  no  way
CC   modified and this statement is not removed.  Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U68168; AAC53206.1; -.
KW   Hydrolase; Pyridoxal phosphate.
FT   BINDING     276    276       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   464 AA;  52453 MW;  37EE19F0 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 464;
Best Local Similarity   75.0%;  Pred. No. 26;
Matches    6;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy      2 RLAIRLDE 9
          |:|:|||
Db     28 RVALRLDE 35

RESULT 6
YMMS_PSEAE
ID   YMMS_PSEAE     STANDARD;      PRT;   464 AA.
AC   P28812;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   HYPOTHETICAL PROTEIN IN MMSB 3'REGION (ORF1) (FRAGMENT).
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE; 92317087.
RA   STEELE M.I., LORENZ D., HATTER K., PARK A., SORATCH J.R.;
RT   "Characterization of the mmsAB operon of Pseudomonas aeruginosa PAO
RT   encoding methylmalonate-semialdehyde dehydrogenase and 3-
RT   hydroxyisobutyrate dehydrogenase.";
RL   J. Biol. Chem. 267:13585-13592(1992).
CC   -!- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
CC       BINDING OF AMP TO THEIR SUBSTRATE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M84911; AAA25893.1; -.
DR   PIR; S27604; S27604.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PFAM; PF00501; AMP-binding; 1.
KW   Hypothetical protein.
FT   NON_TER      464    464
SQ   SEQUENCE   464 AA;  51208 MW;  FE491D7C CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 464;
Best Local Similarity   55.6%;  Pred. No. 26;
Matches    5;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    0;

Qy      1 YRLAIRLDE 9
            | ||:|:|:
Db    146 YELALRIDD 154

RESULT 7
THS_PYRKO
ID   THS_PYRKO     STANDARD;     PRT;    546 AA.
AC   Q52500;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   THERMOSOME SUBUNIT (HEAT-SHOCK PROTEIN).
GN   THS.
OS   Pyrococcus kodakaraensis.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOD1;
RA   IZAWA Y., KAKIHARA H., TAKAGI M., IMANAKA T.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
CC       VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).
CC   -!- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D29672; BAA06143.1; -.
DR   HSSP; P48424; 1ASX.
DR   PROSITE; PS00750; TCP1_1; 1.
DR   PROSITE; PS00751; TCP1_2; 1.
DR   PROSITE; PS00995; TCP1_3; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
```

```
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   546 AA;  59158 MW;  5B3C9283 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 546;
Best Local Similarity   100.0%;  Pred. No. 31;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 LAIRLDE 9
            |||||||
Db    418 LAIRLDE 424

RESULT 8
THSA_THEK1
ID   THSA_THEK1    STANDARD;     PRT;    548 AA.
AC   O24729;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   THERMOSOME, ALPHA SUBUNIT (CHAPERONIN ALPHA SUBUNIT).
GN   THSA.
OS   Thermococcus sp. (strain KS-1).
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98022908.
RA   YOSHIDA T., YOHDA M., IIDA T., MARUYAMA T., TAGUCHI H., YAZAKI K.,
RA   OHTA T., ODAKA M., ENDO I., KAGAWA Y.;
RT   "Structural and functional characterization of homo-oligomeric
RT   complexes of alpha and beta chaperonin subunits from the
RT   hyperthermophilic archaeum Thermococcus strain KS-1.";
RL   J. Mol. Biol. 273:635-645(1997).
CC   -!- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
CC       VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).
CC   -!- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB001080; BAA22207.1; -.
DR   HSSP; P48424; 1ASX.
DR   PROSITE; PS00750; TCP1_1; 1.
DR   PROSITE; PS00751; TCP1_2; 1.
DR   PROSITE; PS00995; TCP1_3; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
KW   Chaperone; ATP-binding; Multigene family.
SQ   SEQUENCE   548 AA;  59191 MW;  08FCFB81 CRC32;

Query Match             65.3%;  Score 32;  DB 1;  Length 548;
Best Local Similarity   100.0%;  Pred. No. 32;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 LAIRLDE 9
            |||||||
Db    418 LAIRLDE 424

RESULT 9
GYRA_RICPR
ID   GYRA_RICPR    STANDARD;     PRT;    905 AA.
AC   P41080;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
```

```
DE   DNA GYRASE SUBUNIT A (EC 5.99.1.3).
GN   GYRA OR RP206.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 95129858.
RA   WOOD D.O., WAITE R.T.;
RT   "Sequence analysis of the Rickettsia prowazekii gyrA gene.";
RL   Gene 151:191-196(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 99039499.
RA   ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA   SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NASLUND A.K.,
RA   ERIKSSON A.-S., WINKLER H.H., KURLAND C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC       STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC       INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC       DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC   -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC       OF DOUBLE-STRANDED DNA.
CC   -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC       BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC       ENZYME FORMS AN A2B2 TETRAMER.
```

```
DR   EMBL; U02931; AAA68146.1; -.
DR   EMBL; AJ235270; CAA14671.1; -.
DR   PFAM; PF00521; DNA_topoisoIV; 1.
KW   Topoisomerase; Isomerase; DNA-binding.
FT   ACT_SITE    123    123    DNA CLEAVAGE (BY SIMILARITY).
SQ   SEQUENCE   905 AA;  101080 MW;  EFBC8ADA CRC32;

Query Match            65.3%;  Score 32;  DB 1;  Length 905;
Best Local Similarity  75.0%;  Pred. No. 54;
Matches   6;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    2  RLAIRLDE  9
         ::||||||
Db  665  KIAIRLDE  672

RESULT 10
VGLC_HSVE4
ID   VGLC_HSVE4    STANDARD;    PRT;    485 AA.
AC   P22596;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCOPROTEIN C PRECURSOR (GLYCOPROTEIN 13).
GN   GC OR GP13.
OS   Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus
OS   type 1 subtype 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91021040.
RA   NICOLSON L., ONIONS D.E.;
RT   "The nucleotide sequence of the equine herpesvirus 4 gC gene
RT   homologue.";
RL   Virology 179:378-387(1990).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
```

```
DR   EMBL; M58031; AAA46083.1; -.
DR   EMBL; A21044; CAA01528.1; -.
DR   PIR; B45343; B45343.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     32
FT   CHAIN        33    485          GLYCOPROTEIN C.
FT   DOMAIN       33    444          EXTRACELLULAR.
FT   TRANSMEM    445    468
FT   CARBOHYD     60     60          POTENTIAL.
FT   CARBOHYD     61     61          POTENTIAL.
FT   CARBOHYD     66     66          POTENTIAL.
FT   CARBOHYD     67     67          POTENTIAL.
FT   CARBOHYD     72     72          POTENTIAL.
FT   CARBOHYD    108    108          POTENTIAL.
FT   CARBOHYD    116    116          POTENTIAL.
FT   CARBOHYD    147    147          POTENTIAL.
FT   CARBOHYD    220    220          POTENTIAL.
FT   CARBOHYD    225    225          POTENTIAL.
FT   CARBOHYD    286    286          POTENTIAL.
SQ   SEQUENCE   485 AA;  52509 MW;  63F72464 CRC32;

Query Match            63.3%;  Score 31;  DB 1;  Length 485;
Best Local Similarity  60.0%;  Pred. No. 45;
Matches   6;  Conservative   2;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1  YRLAIRLDER  10
         ||||:|::|
Db  120  YRLEIHLNQR  129

RESULT 11
MURF_MYCTU
ID   MURF_MYCTU    STANDARD;    PRT;    510 AA.
AC   O06220;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UDP-N-ACETYLMURAMOYLALANYL-D-GLUTAMYL-2,6-DIAMINOPIMELATE--D-ALANYL-D-
DE   ALANYL LIGASE (EC 6.3.2.15) (UDP-MURNAC-PENTAPEPTIDE SYNTHETASE)
DE   (D-ALANYL-D-ALANINE-ADDING ENZYME).
GN   MURF OR RV2157C OR MTCY270.11.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON J., SQUARES S., SQUARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
```

```
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC       STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC       PRECURSOR OF MUREIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL
CC       -MESO-2,6-DIAMINOHEPTANEDIOATE + D-ALANYL-D-ALANYL = ADP +
CC       ORTHOPHOSPHATE + UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GAMMA-GLUTAMYL-6-
CC       CARBOXY-L-LYSYL-D-ALANYL-D-ALANINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   -------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------
DR   EMBL; Z95388; CAB08670.1; -.
DR   PFAM; PF01225; Mur_ligase; 1.
KW   Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW   ATP-binding.
FT   NP_BIND     136    142      ATP (POTENTIAL).
SQ   SEQUENCE   510 AA;  51632 MW;  4F25A40A CRC32;

Query Match           63.3%;  Score 31;  DB 1;  Length 510;
Best Local Similarity 85.7%;  Pred. No. 47;
Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     2 RLAIRLD 8
         |||:|||
Db   412 RLAVRLD 418

RESULT 12
ESTP_DROME
ID   ESTP_DROME     STANDARD;      PRT;   544 AA.
AC   P18167;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   ESTERASE P PRECURSOR (EC 3.1.1.1) (EST-P) (CARBOXYLIC-ESTER
DE   HYDROLASE).
GN   ESTP OR EST-P.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 90136038.
RA   COLLET C., NIELSEN K.M., RUSSELL R.J., KARL M., OAKESHOTT J.G.,
RA   RICHMOND R.C.;
RT   "Molecular analysis of duplicated esterase genes in Drosophila
RT   melanogaster.";
RL   Mol. Biol. Evol. 7:9-28(1990).
CC   -!- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O = AN ALCOHOL
CC       + A CARBOXYLIC ANION.
CC   -!- SUBUNIT: MONOMER.
CC   -!- DEVELOPMENTAL STAGE: MAINLY IN LATE LARVAE.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
CC   -------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------
DR   EMBL; M33780; AAA28520.1; -.
DR   PIR; B34089; B34089.
HSSP; P21836; 1MAH.
DR   FLYBASE; FBgn0000594; Est-P.
DR   PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR   PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
DR   PFAM; PF00135; COesterase; 1.
KW   Hydrolase; Serine esterase; Glycoprotein; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    544      ESTERASE P.
FT   ACT_SITE    206    206      BY SIMILARITY.
FT   ACT_SITE    466    466      BY SIMILARITY.
FT   DISULFID     83    102      BY SIMILARITY.
FT   DISULFID    258    270      BY SIMILARITY.
FT   DISULFID    514    535      POTENTIAL.
FT   CARBOHYD     75     75      POTENTIAL.
FT   CARBOHYD    114    114      POTENTIAL.
FT   CARBOHYD    262    262      POTENTIAL.
FT   CARBOHYD    456    456      POTENTIAL.
SQ   SEQUENCE   544 AA;  61230 MW;  E9F6EEDD CRC32;

Query Match           63.3%;  Score 31;  DB 1;  Length 544;
Best Local Similarity 66.7%;  Pred. No. 51;
Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     1 YRLAIRLDE 9
         ||:|| ||
Db   479 YRIGIRPDE 487

RESULT 13
RPC3_YEAST
ID   RPC3_YEAST     STANDARD;      PRT;   654 AA.
AC   P32349; Q06591;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE III 74 KD POLYPEPTIDE (EC 2.7.7.6) (C74).
GN   RPC3 OR RPC82 OR YPR190C OR P9677.11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 93024385.
RA   CHIANNILKULCHAI N., STALDER R., RIVA M., CARLES C., WERNER M.,
RA   SENTENAC A.;
RT   "RPC82 encodes the highly conserved, third-largest subunit of RNA
RT   polymerase C (III) from Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 12:4433-4440(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA   FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
RA   HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
RA   JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
RA   MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
RA   TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,
RA   WILSON R., WATERSTON R.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
CC       SUBUNITS. THIS SUBUNIT IS THE THIRD LARGEST COMPONENT OF RNA
CC       POLYMERASE III.
```

```
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X63500; CAA45072.1; -.
DR   EMBL; U25841; AAB64619.1; -.
DR   PIR; S31298; S31298.
DR   SGD; L0001693; RPC82.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW   Nuclear protein.
FT   DOMAIN      581    602       LEUCINE-ZIPPER.
FT   CONFLICT    637    637       V -> L (IN REF. 1).
SQ   SEQUENCE   654 AA;  74016 MW;  9E17F4F8 CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 654;
Best Local Similarity   50.0%;  Pred. No. 62;
Matches    5;  Conservative    5;  Mismatches    4;  Indels    1;  Gaps    0;

Qy      1 YRLAIRLDER 10
          |:::|:||:|:
Db    312 YKIALRLTEQ 321


RESULT 14
ID   YM37_YEAST     STANDARD;      PRT;    705 AA.
AC   Q03824;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 81.5 KD PROTEIN IN HLJ1-SMP2 INTERGENIC REGION.
GN   YMR163C OR YM8520.12C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   HUNT S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z49705; CAA89799.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   705 AA;  81466 MW;  6E07A99F CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 705;
Best Local Similarity   55.6%;  Pred. No. 67;
Matches    5;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    0;

Qy      1 YRLAIRLDE 9
          ||::: ||:
Db    136 YRLSLHLDQ 144


RESULT 15
MMLB_MYCLE
ID   MMLB_MYCLE     STANDARD;      PRT;   1014 AA.
AC   O06079;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN MMPL11.
GN   MMPL11 OR MLCL622.16C.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HAMLIN N., CHURCHER C.M., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z95398; CAB08803.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM    188    208       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    235    255       POTENTIAL.
FT   TRANSMEM    279    299       POTENTIAL.
FT   TRANSMEM    311    331       POTENTIAL.
FT   TRANSMEM    373    393       POTENTIAL.
FT   TRANSMEM    409    429       POTENTIAL.
FT   TRANSMEM    530    550       POTENTIAL.
FT   TRANSMEM    560    580       POTENTIAL.
FT   TRANSMEM    598    618       POTENTIAL.
FT   TRANSMEM    649    669       POTENTIAL.
FT   TRANSMEM    671    691       POTENTIAL.
SQ   SEQUENCE   1014 AA;  109875 MW;  A2FC256A CRC32;

Query Match             63.3%;  Score 31;  DB 1;  Length 1014;
Best Local Similarity   50.0%;  Pred. NO. 99;
Matches    5;  Conservative    4;  Mismatches    4;  Indels    1;  Gaps    0;

Qy      1 YRLAIRLDER 10
          |.:::||:|:|
Db    117 YGVSLRLDDR 126


Search completed: February  8, 2000, 00:59:53
Job time: 3782 sec
```

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:39 ; Search time 209.03 Seconds
                                               (without alignments)
                                               3.317 Million cell updates/sec

Title: US-08-653-294-13
Perfect score: 49
Sequence: 1 YRLAIRLDER 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SPTREMBL_12:*
                 1: sp_archea:*
                 2: sp_bacteria.:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	77.6	1324	2	Q44103	Q44103 amycolatops
2	34	69.4	229	5	Q19415	Q19415 caenorhabdi
3	34	69.4	1154	2	O01375	O01375 neurospora
4	34	69.4	1154	3	O01379	O01379 neurospora
5	33	67.3	264	2	O32734	O32734 agrobacteri
6	33	67.3	554	2	O06421	O06421 mycobacteri
7	32	65.3	151	1	O58407	O58407 pyrococcus
8	32	65.3	185	3	O55192	O55192 synechocyst
9	32	65.3	192	2	O13610	O13610 schizosacch
10	32	65.3	304	5	Q9XU51	Q9XU51 caenorhabdi
11	32	65.3	319	6	Q28004	Q28004 bos taurus
12	32	65.3	349	2	Q9ZH35	Q9ZH35 enterobacte
13	32	65.3	422	2	Q9WXB4	Q9WXB4 acidiphiliu
14	32	65.3	548	1	Q9Y8I3	Q9Y8I3 pyrococcus
15	32	65.3	1044	3	Q00943	Q00943 pichia angu
16	31	63.3	236	2	Q9X0G1	Q9X0G1 thermotoga
17	31	63.3	246	2	O54045	O54045 pseudomonas
18	31	63.3	277	2	O07463	O07463 rhodopseudo
19	31	63.3	302	5	Q25608	Q25608 onchocerca
20	31	63.3	356	10	Q9XFP1	Q9XFP1 arabidopsis
21	31	63.3	456	5	Q9XX98	Q9XX98 caenorhabdi
22	31	63.3	463	1	Q9YAG3	Q9YAG3 aeropyrum p
23	31	63.3	485	12	O39258	O39258 equine herp
24	31	63.3	517	2	O69556	O69556 mycobacteri
25	31	63.3	573	4	O94830	O94830 homo sapien
26	31	63.3	740	10	O82436	O82436 cucumis mel
27	31	63.3	741	10	O81122	O81122 malus domes
28	31	63.3	845	5	O01914	O01914 caenorhabdi
29	31	63.3	1110	3	Q92198	Q92198 aspergillus
30	31	63.3	1254	13	Q9YHU2	Q9YHU2 brachydanio
31	31	63.3	1524	1	Q54436	Q54436 staphyloche
32	31	63.3	1699	5	O44344	O44344 strongyloce
33	31	63.3	2282	11	Q61479	Q61479 mus musculu
34	30	61.2	35	12	Q65737	Q65737 bluetongue
35	30	61.2	124	2	O05227	O05227 bacillus su
36	30	61.2	125	1	Q9YBF7	Q9YBF7 aeropyrum p
37	30	61.2	128	2	O33420	O33420 pseudomonas
38	30	61.2	130	2	O05281	O05281 escherichia
39	30	61.2	135	2	O52150	O52150 escherichia
40	30	61.2	135	2	O85624	O85624 escherichia
41	30	61.2	141	2	O67689	O67689 aquifex aeo
42	30	61.2	199	2	Q27010	Q27010 toxoplasma
43	30	61.2	284	5	Q23366	Q23366 caenorhabdi
44	30	61.2	312	5	Q17557	Q17557 caenorhabdi
45	30	61.2	318	5	O44027	O44027 toxoplasma

ALIGNMENTS

RESULT   1
Q44103
ID   Q44103;   PRELIMINARY;   PRT;   1324 AA.
AC   Q44103;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE   PEPTIDE-SYNTHETASE (FRAGMENT).
GN   APS.
OS   Amycolatopsis mediterranei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 5908;
RX   MEDLINE; 97449857.
RA   PELZER S., REICHERT W., HUPPERT M., HECKMANN D., WOHLLEBEN W.;
RT   "Cloning and analysis of a peptide synthetase gene of the balhimycin
RT   producer Amycolatopsis mediterranei DSM5908 and development of a gene
RT   disruption/replacement system.";
RL   J. Biotechnol. 56:115-128(1997).;
DR   EMBL; X97860; CAA66454.1; -.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PFAM; PF00501; AMP-binding; 1.
DR   PFAM; PF00668; DUF4; 2.
DR   PFAM; PF00550; PP-binding; 1.
KW   Ligase.
FT   NON_TER   1   1
FT   NON_TER   1324   1324
SQ   SEQUENCE   1324 AA;   142666 MW; 2C08588E CRC32;

Query Match   77.6%; Score 38; DB 2; Length 1324;
Best Local Similarity   80.0%;   Pred. No. 20;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIRLDER 10
            ||:| ||||
Db   969 YRVAGRLDER 978

RESULT   2
Q19415

```
ID   Q19415      PRELIMINARY;      PRT;    229 AA.
AC   Q19415;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F13E9.10 PROTEIN.
GN   F13E9.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MCMURRAY A.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z69383; CAA93413.1; -.
SQ   SEQUENCE   229 AA;  26620 MW;  F822FE98 CRC32;

Query Match             69.4%;
Best Local Similarity   70.0%;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    2;  Gaps   0;

Qy    1 YRLAIRLDER 10
        | |||||:|
Db  115 YEQAIRLDKR 124

RESULT 3
Q01375
ID   Q01375      PRELIMINARY;      PRT;    1154 AA.
AC   Q01375;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 130.4 KD PROTEIN.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J1518;
RX   MEDLINE; 94203179.
RA   CAMBARERI E.B., HELBER J., KINSEY J.A.;
RT   "Tad1-1, an active LINE-like element of Neurospora crassa.";
RL   Mol. Gen. Genet. 242:658-665(1994).
DR   EMBL; L25662; AAA21781.1; -.
DR   PFAM; PF00078; rvt; 1.
KW   Hypothetical protein.
FT   DOMAIN     1019   1022      POLY-LYS.
FT   DOMAIN     1029   1034      POLY-GLU.
SQ   SEQUENCE   1154 AA;  130398 MW;  DF0BA680 CRC32;

Query Match             69.4%;
Best Local Similarity   66.7%;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps   0;

Qy    1 YRLAIRLDE 9
        |||| : |:|
Db  1136 YRLAVELEE 1144

RESULT 4
Q01379
ID   Q01379      PRELIMINARY;      PRT;    1154 AA.
AC   Q01379;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 130.5 KD PROTEIN.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J1518;
RX   MEDLINE; 94203179.
RA   CAMBARERI E.B., HELBER J., KINSEY J.A.;
RT   "Tad1-1, an active LINE-like element of Neurospora crassa.";
RL   Mol. Gen. Genet. 242:658-665(1994).
DR   EMBL; L25663; AAA21792.1; -.
DR   PFAM; PF00078; rvt; 1.
KW   Hypothetical protein.
FT   DOMAIN     1019   1022      POLY-LYS.
FT   DOMAIN     1029   1034      POLY-GLU.
SQ   SEQUENCE   1154 AA;  130470 MW;  7FBE8EAF CRC32;

Query Match             69.4%;        Score 34;  DB 3;  Length 1154;
Best Local Similarity   66.7%;        Pred. No. 1.2e+02;
Matches    6;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps   0;

Qy    1 YRLAIRLDE 9
        |||| : |:|
Db  1136 YRLAVELEE 1144

RESULT 5
O32734
ID   O32734      PRELIMINARY;      PRT;    264 AA.
AC   O32734;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ATTE PROTEIN.
GN   ATTE.
OS   Agrobacterium tumefaciens.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Agrobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C58;
RX   MEDLINE; 96359388.
RA   MATHYSSE A.G., YARNALL H.A., YOUNG N.;
RT   "Requirement for genes with homology to ABC transport systems for
RT   attachment and virulence of Agrobacterium tumefaciens.";
RL   J. Bacteriol. 178:5302-5308(1996).
DR   EMBL; U59485; AAB67299.1; -.
DR   PFAM; PF00005; ABC_tran; 1.
SQ   SEQUENCE   264 AA;  28745 MW;  D5629761 CRC32;

Query Match             67.3%;        Score 33;  DB 2;  Length 264;
Best Local Similarity   66.7%;        Pred. No. 40;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps   0;

Qy    2 RLAIRLDER 10
        ||| || | |
Db  246 RLAVKLDRR 254
```

```
RESULT   6
006421
ID   006421      PRELIMINARY;       PRT;    554 AA.
AC   006421;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MEND.
GN   MEND.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV.
RA   BROWN D., CHURCHER C.M.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV.
RA   BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV.
RX   MEDLINE: 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; Z95558; CAB08956.1; -.
SQ   SEQUENCE  554 AA; 57835 MW;  C42C89FC CRC32;

Query Match              67.3%;
Best Local Similarity   66.7%;
Matches    6;  Conservative    2; Mismatches    2; Indels    1; Gaps   0;

Qy      2 RLAIRLDER 10
          || |:||||
Db     48 RLHVRIDER 56

RESULT   7
058407
ID   058407      PRELIMINARY;       PRT;    151 AA.
AC   058407;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   151AA LONG HYPOTHETICAL FRXA PROTEIN.
GN   PH0674.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE: 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000003; BAA29765.1; -.
SQ   SEQUENCE  151 AA; 17160 MW;  11AACD59 CRC32;

Query Match              65.3%;  Score 32;  DB 1;  Length 151;
Best Local Similarity   77.8%;  Pred. No. 36;
Matches    7;  Conservative    0; Mismatches    0; Indels    2; Gaps   0;

Qy      2 RLAIRLDER 10
          || |||||
Db     96 RLLIELDER 104

RESULT   8
Q55192
ID   Q55192      PRELIMINARY;       PRT;    185 AA.
AC   Q55192;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   HYPOTHETICAL 20.8 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE: 96127529.
RA   KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA   SUGIURA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE: 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D64001; BAA10334.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE  185 AA; 20830 MW;  365A078D CRC32;

Query Match              65.3%;  Score 32;  DB 2;  Length 185;
Best Local Similarity   60.0%;  Pred. No. 44;
Matches    6;  Conservative    6; Mismatches    3; Indels    1; Gaps   0;

Qy      1 YRLAIRLDER 10
          |||:|:.:|
Db     35 YRLALRILQR 44

RESULT   9
O13610
ID   O13610      PRELIMINARY;       PRT;    192 AA.
AC   O13610;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   HYPOTHETICAL 22.4 KD PROTEIN.
GN   PI019.
OS   Schizosaccharomyces pombe (Fission yeast).
```

```
OC   Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972 H;
RA   KUSHIDA N., YAMAZAKI S., TANAKA T., JINNO K., HAIKAWA Y., YAMAZAKI J.,
RA   YAMAMOTO S., SEKINE M., OGUCHI A., NAGAI Y., SAKAI M., AOKI K.,
RA   OGURA K., OTSUKA R., KUDOH Y., YANAGIDA M., MACHIDA M., ZHANG M.Q.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB004533; BAA21398.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   192 AA;  22352 MW;  F43F0759 CRC32;

Query Match             65.3%;  Score 32;  DB 3;  Length 192;
Best Local Similarity   58.3%;  Pred. No. 46;
Matches    7;  Conservative    3;  Mismatches    0;  Indels    2;  Gaps    1;

Qy        1 YRLAIRL--DER 10
            |:|:|: |||
Db      115 YRIALRIAKDER 126

RESULT 10
Q9XU51
ID   Q9XU51          PRELIMINARY;          PRT;   304 AA.
AC   Q9XU51;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T08G3.6 PROTEIN.
GN   T08G3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   LLOYD C.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z83238; CAB05796.1; -.
SQ   SEQUENCE   304 AA;  34207 MW;  F9701C2D CRC32;

Query Match             65.3%;  Score 32;  DB 5;  Length 304;
Best Local Similarity   77.8%;  Pred. No. 75;
Matches    7;  Conservative    7;  Mismatches    0;  Indels    2;  Gaps    0;

Qy        1 YRLAIRLDE 9
            ||||| ||
Db      287 YRLAINNDE 295

RESULT 11
```

```
Q28004
ID   Q28004          PRELIMINARY;          PRT;   319 AA.
AC   Q28004;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   11-CIS-RETINOL DEHYDROGENASE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95386398.
RA   DRIESSEN C.A., JANSSEN B.P., WINKENS H.J., VAN VUGT A.H., LEEUW T.L.,
RA   JANSSEN J.J.;
RT   "Cloning and expression of a cDNA encoding bovine retinal pigment
RT   epithelial 11-cis retinol dehydrogenase.";
RL   Invest. Ophthalmol. Vis. Sci. 36:1988-1996(1995).
DR   EMBL; L36533; AAA80694.1; -.
DR   HSSP; P14061; 1FDW.
DR   PFAM; PF00106; adh_short; 1.
DR   PRINTS; PR00080; ALCDHDRGNASE.
SQ   SEQUENCE   319 AA;  34400 MW;  37A78DAA CRC32;

Query Match             65.3%;  Score 32;  DB 6;  Length 319;
Best Local Similarity   75.0%;  Pred. No. 79;
Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 LAIRLDER 10
            ||:|||:|
Db       45 LALRLDQR 52

RESULT 12
Q9ZH35
ID   Q9ZH35          PRELIMINARY;          PRT;   349 AA.
AC   Q9ZH35;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   NITROGEN REGULATORY PROTEIN.
GN   NTRB.
OS   Enterobacter gergoviae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Enterobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=57-7;
RA   DONG Y.M., LI J.D.;
RT   "The cloning of glnA, ntrB, and ntrC from Enterobacter gergoviae 57-7
RT   and their characterization.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF072440; AAC69321.1; -.
SQ   SEQUENCE   349 AA;  38412 MW;  A9F4BA43 CRC32;

Query Match             65.3%;  Score 32;  DB 2;  Length 349;
Best Local Similarity   75.0%;  Pred. No. 86;
Matches    6;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    0;

Qy        1 YRLAIRLD 8
            |||| |:|
Db      277 YRLAARID 284

RESULT 13
Q9WXB4
ID   Q9WXB4          PRELIMINARY;          PRT;   422 AA.
AC   Q9WXB4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
```

DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   LIGHT-INDEPENDENT PROTOCHLOROPHYLLIDE REDUCTASE.
GN   BCHN.
OS   Acidiphilium rubrum.
OC   Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC   Acidiphilium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MASUDA T., INOUE K., MASUDA M., NAGAYAMA M., OHTA H., SHIMADA H.,
RA   TAKAMIYA K.;
RT   "The metal-insertion step of bacteriochlorophyll biosynthesis in an
RT   aerobic bacterium Acidiphilium rubrum, which produces zinc-containing
RT   bacteriochlorophyll as natural photosynthetic pigment.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB017351; BAA76536.1; -.
SQ   SEQUENCE   422 AA;   45864 MW;   4248EA89 CRC32;

Query Match   65.3%; Score 32; DB 2; Length 422;
Best Local Similarity   77.8%; Pred. No. 1.1e+02;
Matches   7; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIRLDER 10
           | | |||||
Db   123 RAAARLDER 131

RESULT 14
Q9Y8I3
ID   Q9Y8I3   PRELIMINARY;   PRT;   548 AA.
AC   Q9Y8I3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CHAPERONIN LIKE PROTEIN ALPHA SUBUNIT.
GN   CPKA.
OS   Pyrococcus kodakaraensis.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOD1;
RX   MEDLINE; 99203147.
RA   IZUMI M., FUJIWARA S., TAKAGI M., KANAYA S., IMANAKA T.;
RT   "Isolation and characterization of a second subunit of molecular
RT   chaperonin from pyrococcus kodakaraensis KOD1: analysis of an ATPase-
RT   deficient mutant enzyme.";
RL   Appl. Environ. Microbiol. 65:1801-1805(1999).
DR   EMBL; AB018432; BAA76952.1; -.
DR   PROSITE; PS00750; TCP1_1; 1.
DR   PROSITE; PS00751; TCP1_2; 1.
DR   PROSITE; PS00995; TCP1_3; 1.
SQ   SEQUENCE   548 AA;   59169 MW;   10DA6C62 CRC32;

Query Match   65.3%; Score 32; DB 1; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   3 LAIRLDE 9
           |||||||
Db   418 LAIRLDE 424

RESULT 15
Q00943
ID   Q00943   PRELIMINARY;   PRT;   1044 AA.
AC   Q00943;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NITRITE REDUCTASE.
GN   YNI1.
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCYC 495;
RA   BRITO N., AVILA J., PEREZ M., GONZALEZ C., SIVERIO J.M.;
RL   J. Biochem. 317:89-95(1996).
DR   EMBL; Z68122; CAA92206.1; -.
DR   PFAM; PF01077; NIR_SIR; 1.
DR   PFAM; PF00355; Rieske; 1.
DR   PRINTS; PR00397; SIROHAEM.
SQ   SEQUENCE   1044 AA;   116574 MW;   59F4D4B1 CRC32;

Query Match   65.3%; Score 32; DB 3; Length 1044;
Best Local Similarity   75.0%; Pred. No. 2.7e+02;
Matches   6; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   3 LAIRLDER 10
           ||:||:||
Db   696 LAVRLEER 703

Search completed: February 8, 2000, 13:17:41
Job time: 32490 sec

OM of: US-08-653-294-13 to: GenEmbl:*   out_format : pfs
Date: Feb 8, 2000 4:39 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15779/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-13
Query length: 10
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 11370.480000

```
gb_pr3:AC005605   -   35.00   88.18   6.9e+03   37640  ! AC005605 Homo sapiens subt
gb_pl1:AB015475   +   35.00   80.58   1.8e+04   84325  ! AB015475 Arabidopsis thall
gb_pl2:ATF17I5    -   35.00   80.58   1.8e+04   86748  ! AL031032 Arabidopsis thall

seq_name: gb_pr2:HS1106N18

seq_documentation_block:
LOCUS       HS1106N18  142336 bp    DNA            PRI    23-NOV-1999
DEFINITION  Human DNA sequence from clone 1106N18 on chromosome 20q13.2-13.2,
            complete sequence.
ACCESSION   AL035457
VERSION     AL035457.13  GI:6143575
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142336)
AUTHORS     Clark,G.
TITLE       Direct Submission
JOURNAL     Submitted (30-OCT-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 29, 1999 this sequence version replaced gi:606875.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep IMPORTANT: This
            sequence is not the entire insert of clone 1106N18. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
            The true left end of clone 906P16 is at 142237 in this sequence.
            The true right end of clone dJ1193N1 is at 79367 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            1106N18 is from the library RPCI-5 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..142336
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="q13.2-13.2"
                     /clone_lib="RPCI-5"
                     /clone="RP5-1106N18"
BASE COUNT    37456 a  31918 c  32886 g  40076 t
ORIGIN

alignment_scores:
                     Quality:    40.00          Length:     10
                     Ratio:       4.000          Gaps:        0
Percent Similarity: 100.000   Percent Identity:  80.000

alignment_block:
US-08-653-294-13 x HS1106N18/rev
```

score_list:

sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_pr2:HS1106N18	-	40.00	96.64	2.3e+03	142336	! AL035457 Human DNA sequence
gb_htg2:AC004965	+	40.00	89.16	6.1e+03	323792	! AC004965 Homo sapiens clone
gb_om:AF064209	-	39.00	126.37	51.38	3452	! AF064209 Macropus eugenii inte
gb_htg1:AC018207	+	39.00	103.02	1.0e+03	44887	! AC018207 Drosophila melanogas
gb_in2:AC005268	+	39.00	96.72	2.3e+03	89791	! AC005268 Drosophila melanogas
gb_bal:AMPEPSYNT	+	38.00	120.97	102.66	3975	! X97860 Amycolatopsis mediterra
gb_pr2:HS204E5	+	38.00	89.24	6.0e+03	129969	! Z98941 Human DNA sequence fr
gb_pr4:AC009239	+	38.00	88.80	6.4e+03	136371	! AC009239 Homo sapiens clone
gb_htg4:AC011830	+	38.00	86.97	8.0e+03	166832	! AC011830 Homo sapiens chromo
gb_htg3:AC010086	+	38.00	85.91	9.2e+03	187447	! AC010086 Homo sapiens clone
gb_htg4:AC008591	+	38.00	84.80	1.1e+04	211769	! AC008591 Homo sapiens chromo
gb_in1:AF026152	+	37.00	135.98	14.97	486	! AF026152 Caenorhabditis elegans
gb_bal:HSMOXMXU	+	37.00	126.87	48.18	1323	! Y09104 Hyphomicrobium sp. DNA,
gb_in1:CELC04F6	+	37.00	100.09	1.5e+03	25083	! U42835 Caenorhabditis elegans
gb_htg5:AC015346	+	37.00	99.72	1.6e+03	26121	! AC015346 Drosophila melanogas
gb_in1:CELF09F7	+	37.00	97.82	2.0e+03	32202	! U00050 Caenorhabditis elegans
gb_htg6:AC011071	+	37.00	84.34	1.1e+04	141615	! AC011071 Drosophila melanoga
gb_htg4:AC008274	+	37.00	83.49	1.3e+04	155524	! AC008274 Homo sapiens clone
gb_pr2:AC002465	+	37.00	83.47	1.3e+04	155881	! AC002465 Human BAC clone RG3
gb_htg6:AC010847	+	37.00	82.44	1.4e+04	174541	! AC010847 Drosophila melanoga
gb_htg3:AC008537	+	37.00	81.80	1.6e+04	187246	! AC008537 Homo sapiens chromo
gb_pr2:HSM800934	+	36.00	112.35	310.04	4150	! AL117426 Homo sapiens mRNA; cD
gb_in1:DMUSHAPE	+	36.00	111.11	363.76	4759	! Y12322 D.melanogaster mRNA for
gb_in1:CEW08G11	+	36.00	94.76	4.0e+03	28690	! Z92817 Caenorhabditis elegans
gb_htg3:AC010400	+	36.00	92.45	4.0e+03	36974	! AC010400 Homo sapiens chromos
gb_in2:AC004368	+	36.00	84.37	1.1e+04	89885	! AC004368 Drosophila melanogas
gb_in2:AC005269	+	36.00	83.02	1.4e+04	104278	! AC005269 Drosophila melanoga
gb_pr3:HS281H8	+	36.00	82.50	1.4e+04	110414	! AL031133 Human DNA sequence
gb_htg1:AP000585	+	36.00	81.48	1.6e+04	123497	! AP000585 Homo sapiens chromo
gb_htg7:AC008002	+	36.00	81.25	1.7e+04	126629	! AP008002 Drosophila melanoga
gb_htg7:AC017155	+	36.00	78.90	2.2e+04	163954	! AC017155 Drosophila melanoga
gb_htg1:AP000642	+	36.00	78.56	2.3e+04	170145	! AP000642 Homo sapiens chromo
gb_ov:AF073712	+	35.00	125.79	55.32	603	! AF073712 Salvelinus fontinalis
gb_pl2:CNS01BRR	+	35.00	124.62	64.30	686	! AL114607 Botrytis cinerea strai
gb_pl2:CNS01A6K	+	35.00	123.45	74.70	780	! AL112548 Botrytis cinerea strai
gb_pl1:ANH3GENE	+	35.00	119.56	123.01	1196	! X55548 A.nidulans gene for cor
gb_pr2:HSA01I741	-	35.00	114.45	236.75	2096	! AJ011741 Homo sapiens TOP2 alp
gb_pl1:ENU12630	-	35.00	112.21	315.82	2683	! U12630 Emericella nidulans R15
gb_in2:CEU58085	-	35.00	111.15	361.60	3013	! U58085 Caenorhabditis elegans
gb_pl1:SCLYP1	-	35.00	109.43	450.72	3639	! X67315 S.cerevisiae LYP1 gene
gb_ba2:AE000541	+	35.00	97.84	2.0e+03	13012	! AE000541 Helicobacter pylori
gb_htg5:AC014153	+	35.00	90.49	5.1e+03	29183	! AC014153 Drosophila melanogas

Align seg 1/1 to reverse of: HS1106N18 from: 1 to: 142336

```
      1 TyrArgLeuAlaIleArgLeuAspGluArg 10
        ||||||||||:::||||||||||::|||
 112966 TATAGACTGTGTATCAGGCTGGACCACAGG 112937
```

seq_name: gb_htg2:AC004965

```
seq_documentation_block:
LOCUS       AC004965   323792 bp    DNA             HTG        12-JUN-1998
DEFINITION  Homo sapiens clone DJ1106H14, *** SEQUENCING IN PROGRESS ***, 42
            unordered pieces.
ACCESSION   AC004965
VERSION     AC004965.1  GI:3213032
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 323792)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 323792)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 42 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1848: contig of 1848 bp in length
            *     1849     1867: gap of unknown length
            *     1868     4168: contig of 2301 bp in length
            *     4169     4187: gap of unknown length
            *     4188     6023: contig of 1836 bp in length
            *     6024     6042: gap of unknown length
            *     6043     7886: contig of 1844 bp in length
            *     7887     7905: gap of unknown length
            *     7906     9644: contig of 1739 bp in length
            *     9645     9663: gap of unknown length
            *     9664    11563: contig of 1900 bp in length
            *    11564    11582: gap of unknown length
            *    11583    13437: contig of 1855 bp in length
            *    13438    13456: gap of unknown length
            *    13457    15087: contig of 1631 bp in length
            *    15088    15106: gap of unknown length
            *    15107    17659: contig of 2553 bp in length
            *    17660    17678: gap of unknown length
            *    17679    19643: contig of 1965 bp in length
            *    19644    19662: gap of unknown length
            *    19663    21474: contig of 1812 bp in length
            *    21475    21493: gap of unknown length
            *    21494    23992: contig of 2499 bp in length
            *    23993    24011: gap of unknown length
            *    24012    27269: contig of 3258 bp in length
            *    27270    27288: gap of unknown length
            *    27289    30064: contig of 2776 bp in length
            *    30065    30083: gap of unknown length
            *    30084    33836: contig of 3753 bp in length
            *    33837    33855: gap of unknown length
            *    33856    36684: contig of 2829 bp in length
            *    36685    36703: gap of unknown length
            *    36704    39501: contig of 2798 bp in length
            *    39502    39520: gap of unknown length
            *    39521    43003: contig of 3483 bp in length
            *    43004    43022: gap of unknown length
            *    43023    46858: contig of 3836 bp in length
            *    46859    46877: gap of unknown length
            *    46878    50622: contig of 3745 bp in length
            *    50623    50641: gap of unknown length
            *    50642    54121: contig of 3480 bp in length
            *    54122    54140: gap of unknown length
            *    54141    58079: contig of 3939 bp in length
            *    58080    58098: gap of unknown length
            *    58099    61695: contig of 3597 bp in length
            *    61696    61714: gap of unknown length
            *    61715    66025: contig of 4311 bp in length
            *    66026    66044: gap of unknown length
            *    66045    67681: contig of 1637 bp in length
            *    67682    67700: gap of unknown length
            *    67701    70508: contig of 2808 bp in length
            *    70509    70527: gap of unknown length
            *    70528    75295: contig of 4768 bp in length
            *    75296    75314: gap of unknown length
            *    75315    80364: contig of 5050 bp in length
            *    80365    80383: gap of unknown length
            *    80384    82047: contig of 1664 bp in length
            *    82048    82066: gap of unknown length
            *    82067    88827: contig of 6761 bp in length
            *    88828    88846: gap of unknown length
            *    88847    98762: contig of 9916 bp in length
            *    98763    98781: gap of unknown length
            *    98782   110167: contig of 11386 bp in length
            *   110168   110186: gap of unknown length
            *   110187   122949: contig of 12763 bp in length
            *   122950   122968: gap of unknown length
            *   122969   140267: contig of 17299 bp in length
            *   140268   140286: gap of unknown length
            *   140287   163837: contig of 23551 bp in length
            *   163838   163856: gap of unknown length
            *   163857   192172: contig of 28316 bp in length
            *   192173   192191: gap of unknown length
            *   192192   228203: contig of 36012 bp in length
            *   228204   228222: gap of unknown length
            *   228223   317448: contig of 89226 bp in length
            *   317449   317466: gap of unknown length
            *   317467   318949: contig of 1483 bp in length
            *   318950   318967: gap of unknown length
            *   318968   320503: contig of 1536 bp in length
            *   320504   320521: gap of unknown length
            *   320522   322060: contig of 1539 bp in length
            *   322061   322078: gap of unknown length
            *   322079   323792: contig of 1714 bp in length.
FEATURES            Location/Qualifiers
     source          1..323792
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ1106H14"
BASE COUNT    81610 a   77790 c   79384 g   84233 t    775 others
ORIGIN
```

alignment_scores:
```
                 Quality:     40.00              Length:       10
                   Ratio:      4.000               Gaps:        0
     Percent Similarity: 100.000    Percent Identity:  80.000
```

alignment_block:
US-08-653-294-13 x AC004965

Align seg 1/1 to: AC004965 from: 1 to: 323792

```
      1 TyrArgLeuAlaIleArgLeuAspGluArg 10
        |||||||||:::|||||||||||||::|||
  75519 TATAGACTGTGTATCAGGCTGGACCACAGG 75548
```

seq_name: gb_om:AF064209

seq_documentation_block:
LOCUS   AF064209   3452 bp   DNA   MAM   05-OCT-1999
DEFINITION Macropus eugenii interleukin-5 (IL5) gene, complete cds.
ACCESSION   AF064209
VERSION   AF064209.1 GI:5006325
KEYWORDS   .
SOURCE   tammar wallaby.
ORGANISM   Macropus eugenii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Metatheria; Diprotodontia; Macropodidae; Macropus.
REFERENCE   1 (bases 1 to 3452)
AUTHORS   Hawken,R.J., Maccarone,P., Toder,R., Marshall Graves,J.A. and
            Maddox,J.F.
TITLE   Isolation and characterization of marsupial IL5 genes
JOURNAL   Immunogenetics 49 (11-12), 942-948 (1999)
MEDLINE   99432005
REFERENCE   2 (bases 1 to 3452)
AUTHORS   Hawken,R.J., Maccarone,P., Toder,R., Marshall Graves,J.A. and
            Maddox,J.F.
TITLE   Direct Submission
JOURNAL   Submitted (07-MAY-1998) Veterinary PathoBiology, University of
            Minnesota, 1988 Fitch Avenue, St. Paul, MN 55108, USA
FEATURES   Location/Qualifiers
     source   1..3452
                     /organism="Macropus eugenii"
                     /db_xref="taxon:9315"
                     /chromosome="1"
     mRNA   join(<34..180,406..438,1492..1620,1706..>1816)
                     /gene="IL5"
                     /product="interleukin-5"
     gene   <34..>1816
                     /gene="IL5"
     CDS   join(34..180,406..438,1492..1620,1706..1816)
                     /gene="IL5"
                     /codon_start=1
                     /product="interleukin-5"
                     /protein_id="AAD37462.1"
                     /db_xref="GI:5006326"
                     /translation="MMKILVCLPLTLYAGCVYGIATGNPVSRLVTETLSLITAHRTL
                     LIGNGTLRISIPDPQNHPLCIEEIFQGIETLKNQTAEENVVEKIFQNLSSLKGYITAK
                     EKQCGGERRRVEQFLDYLEEFLRTNIEWNTEWTVES"
BASE COUNT   1046 a   585 c   616 g   1205 t
ORIGIN

alignment_scores:
       Quality: 39.00   Length: 10
       Ratio: 4.333   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x AF064209/rev   ..
Align seg 1/1 to reverse of: AF064209 from: 1 to: 3452

       1 TyrArgAlaIleArgAspGluArg 10
         ||||||::::||| |||::::||||||||
    3345 TACAGAATGGCAACTAGAGTAGATGAAAGG 3316

seq_name: gb_htg7:AC018207

seq_documentation_block:
LOCUS   AC018207   44887 bp   DNA   HTG   09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018207
VERSION   AC018207.1 GI:6552984
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE   1 (bases 1 to 44887)
AUTHORS   Adams,M. and Venter,J.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10213855 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES   Location/Qualifiers
     source   1..44887
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT   12412 a   9869 c   9763 g   12843 t
ORIGIN

alignment_scores:
       Quality: 39.00   Length: 9
       Ratio: 4.875   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 88.889

alignment_block:
US-08-653-294-13 x AC018207/rev   ..
Align seg 1/1 to reverse of: AC018207 from: 1 to: 44887

       1 TyrArgLeuLeuArgLeuAspGlu 9
         ||||||||||||||| |||||
   17401 TATAGACTGCTATCAGGTGCGATGAG 17375

seq_name: gb_in2:AC005268

seq_documentation_block:
LOCUS   AC005268   89791 bp   DNA   INV   07-JUL-1998
DEFINITION Drosophila melanogaster DNA sequence (P1s DS03550 (D225) and
            DS02397 (D206)), complete sequence.
ACCESSION   AC005268 AC004330 AC004314 AC003599
VERSION   AC005268.1 GI:3293207
KEYWORDS   HTG.
SOURCE   Drosophila melanogaster (Subclones in sac from P1 clones DS03550
            (D225) and DS02397 (D206)) DNA.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 89791)
AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Humnasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
TITLE   Sequencing of Drosophila chromosome, region 57B6-57C1
JOURNAL   Unpublished (1997)
REFERENCE   2 (bases 1 to 89791)
AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Humnasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE   Direct Submission
JOURNAL   Submitted (07-JUL-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT   Sequence submitted by:

```
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library locations: 137-37, 136-25.
FEATURES             Location/Qualifiers
     source          1..89791
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="57B6-57C1"
                     /clone="Pls DS03550 (D225) and DS02397 (D206)"
                     /note="DS03550 (D225) extends from bp 1 to bp 16,924 and
                     DS02397 (D206) extends from bp 13,480 to bp 89,791."
BASE COUNT     24290 a  20016 c  20807 g  24678 t
ORIGIN

alignment_scores:
     Quality:              39.00          Length:          9
     Ratio:                4.875          Gaps:            0
     Percent Similarity:   88.889         Percent Identity: 88.889

alignment_block:
US-08-653-294-13 x AC005268

Align seq 1/1  to: AC005268    from: 1   to: 89791

     1 TyrArgLeuAlaIleArgLeuAspGlu 9
       |||||||||||||||   |||||
 24440 TATAGACTGGCTATCAGGTCGCGATGAG 24466

seq_name: gb_bal:AMPEPSYNT

seq_documentation_block:
LOCUS       AMPEPSYNT      3975 bp    DNA             BCT       14-OCT-1997
DEFINITION  Amycolatopsis mediterranei aps gene, partial.
ACCESSION   X97860
VERSION     X97860.1  GI:2576250
KEYWORDS    APS gene; peptide synthetase.
SOURCE      Amycolatopsis mediterranei
ORGANISM    Amycolatopsis mediterranei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Amycolatopsis.
REFERENCE   1  (bases 1 to 3975)
AUTHORS     Pelzer,S., Reichert,W., Huppert,M., Heckmann,D. and Wohlleben,W.
TITLE       Cloning and analysis of a peptide synthetase gene of the balhimycin
            producer Amycolatopsis mediterranei DSM5908 and development of a
            gene disruption/replacement system
JOURNAL     J. Biotechnol. 56 (2), 115-128 (1997)
MEDLINE     97449857
REFERENCE   2  (bases 1 to 3975)
AUTHORS     Pelzer,S.
TITLE       Direct Submission
JOURNAL     Submitted (15-MAY-1996) S. Pelzer, Universitaet Tuebingen,
            Lehrstuhl Mikrobiologie-Biotechnologie, Auf der Morgenstelle 28, D-
            72076 Tuebingen, FRG
REMARK      Revised by author 14-OCT-97
COMMENT     On Oct 30, 1997 this sequence version replaced gi:1483198.
FEATURES             Location/Qualifiers
     source          1..3975
                     /organism="Amycolatopsis mediterranei"
                     /strain="DSM 5908"
                     /db_xref="taxon:33910"
     gene            1..3975
                     /gene="aps"
     CDS             <1..>3975
                     /gene="aps"
                     /codon_start=3
                     /transl_table=11
                     /product="peptide-synthetase"
                     /protein_id="CAA66454.1"
                     /db_xref="GI:2576251"
                     /db_xref="SPTREMBL:Q44103"
                     /translation="IPLDRGEPELTAGFDLTTGPLLRATLVDATTLVLVVHHIVWDGW
SAGVFERELAEFYTAAVEGREPVLPELITQYAGYAAEAVVHDDQLAYWREQLQGAPGR
LAVPTDREAPAQQDFAGDTREFALPAGLAARIGELAAEEDATPFVVQLAFAALLNRY
TGAADLVVGTPVTTRDRPELADLIGYFVNILPRLKIDRAASFRDLVEHVRDTAFDAY
ACLDVPFDVVVDALALERTPRHAPLVQVVFGAHAEDPAPLRFGPLTATRRVHHNGTSK
FDFTWSTFDDGELRGEVEYRTSLFDAATVDRMTHWRALLTAVLSEPDSPLWKIDLEP
VWPVAVTASQPRCLHESEPDSVDRFPGRPAVTYAELDRRANRLAHALLIEAG
VRPGDRVGLLLDRIDAIVVAILAVLKAGAAYVPVDPAAPDDRAAFVFGDTGVRLVVTD
QGTDGPWPVFDLARDVSAYSADRPRVPVRPGDLAYLIFTSGSTGRPKGVAVAHEHAGR
LMASGHAHFGFTETDVWTLFHSVAFDWTVWELMGPLHHGGRLVLVPYLVSRSPEAFTA
LLADEGVTMLCQTPSALRQLETELRTTPRALPALRQVMLGGEALDPAVVRRWFAHASS
APLCNLYGITETTVHVTTHDVPGAFGFERSLIGTPLPHLSAHVLDEWLRPCPAGVPGE
LYIGGGALAHGYWGRAGLTAQRFLPDPFSPVPGARLYRTGDVARRLAGGGLEVVGRCD
SQVKIRGFRIELGEIEHALGAHPAVGACAVTVHDDRLAAYVTGGPDHAELRAHLAKSL
PEHMIPATVTVLDRLPVTVNGKLDRAALPAPAPRAANAYTAPSTPGERLLTELWSDVL
GVPGAGVHDNFFHLGGDSIRAVHLAGKLRDRGWTFTLPDLFGAPTPAALAPLLKPCAG
ETPASRPFAGLSEKDLAKLPGDVVDAYPMAAMQLGMIYHMELSGDAGGYHNVNSYRVA
GRLDERALRAAVAGAIARHPVLRTTFDVIGYRQPMQLVHAEAPAPVETADLRGLSEAA
QRDAVTEVFDGLCAVRFDLRTPPLFRVVAQRLADDVFQLTIAEHHSILDGWSFTSLLT
EILERHADPDAPPAPPASTFRDFVAAEQAVASAESERFWRDRLTGANGALWSSGTD
SAAATAEIPRTLERVLPDPAQLAAIAAAGYPLAAIAHGAHVRALARITGRDRVTTG
LSVNGRLEERSGTEAYGLFLNTVPLVVDCTERDLVRSVHEDEVALLPHRRVPFARLAR
LMAGPRLEACFAFLRFHAIGRLAGSATSIVDDRIGCEPDMRYEPTNFALGVALVQDPA
SGRI"

BASE COUNT      504 a   1504 c   1425 g    542 t
ORIGIN

alignment_scores:
     Quality:              38.00          Length:          10
     Ratio:                4.222          Gaps:            0
     Percent Similarity:   90.000         Percent Identity: 80.000

alignment_block:
US-08-653-294-13 x AMPEPSYNT

Align seq 1/1  to: AMPEPSYNT    from: 1   to: 3975

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       |||||::::|||  ||||||||||||||
  2907 TACCGGGTCGCCGGCCGCCCTCGACGAACGC 2936

seq_name: gb_pr2:HS204E5

seq_documentation_block:
LOCUS       HS204E5      129969 bp    DNA              PRI       22-NOV-1999
DEFINITION  Human DNA sequence from PAC 204E5 on chromosome 12. Contains exon
            similar to Wilms' tumour-related protein QM-like P2x-like receptor,
            ATP ligand gated ion channel, ESTs, CpG island.
ACCESSION   Z98941
VERSION     Z98941.1  GI:2370071
KEYWORDS    12; ATP ligand gated ion channel; CpG island; P2X-like receptor;
            QM-like; Tumour-related protein.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129969)
AUTHORS     Kershaw,J.
TITLE       Direct Submission
JOURNAL     Submitted (29-AUG-1997) Chromosome 12 Project Group
            (http://www.sanger.ac.uk/HGP/Chr12/) Sanger Centre, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is the entire insert of clone 204E5.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together
```

with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of human chromosome 12, constructed by the Sanger Centre chromosome 12 mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr12/

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 204E5 is at 1 in this sequence. The true right end of clone 204E5 is at 129969.

204E5 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.

For further details see http://bacpac.med.buffalo.edu/.

```
FEATURES             Location/Qualifiers
     source          1..129969
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12"
                     /clone_lib="RPCI1"
                     /clone="XX-204E5"
     repeat_region   1..53
                     /note="FLAM_C repeat: matches 59. .111 of consensus"
     repeat_region   203..493
                     /note="AluSg repeat: matches 300. .3 of consensus"
     repeat_region   744..931
                     /note="AluJo repeat: matches 118. .302 of consensus;
                     incomplete repeat"
     repeat_region   1116..1254
                     /note="MIR2 repeat: matches 145. .2 of consensus"
     misc_feature    1319..1383
                     /note="92N15 has 2 versions, both differ from 204E5"
     repeat_region   1331..1374
     repeat_region   1473..1722
                     /note="22 copies of 2 mer 89 % conserved"
     repeat_region   1928..2227
                     /note="AluY repeat: matches 38. .301 of consensus;
                     incomplete repeat"
     repeat_region   3308..3606
                     /note="AluSq repeat: matches 300. .1 of consensus"
     variation       3586..3607
                     /note="AluJb repeat: matches 5. .302 of consensus"
                     /note="clone 92N15; CAAAAAAAAAAAAAAAAAAAG in this entry;
                     deletion"
                     /replace="caaaaaaaaaaaaaaaaaaaag"
     repeat_region   4200..4346
                     /note="AluJo repeat: matches 282. .134 of consensus;
                     incomplete repeat"
     variation       5211..5213
                     /note="clone 92N15; ATG in this entry; substitution"
                     /replace="agg"
     misc_feature    5318..5401
                     /note="differs from 92N15"
     repeat_region   5460..5750
                     /note="AluJb repeat: matches 302. .21 of consensus;
                     incomplete repeat"
     repeat_region   5813..5928
                     /note="AluJb repeat: matches 180. .295 of consensus;
                     incomplete repeat"
     repeat_region   6157..6459
                     /note="AluSx repeat: matches 1. .302 of consensus"
     misc_feature    6761..7142
                     /note="differs from 92N15"
     repeat_region   7138..7438
                     /note="AluSq repeat: matches 300. .1 of consensus"
     repeat_region   7593..7839
                     /note="PABL repeat: matches 10. .273 of consensus"
     repeat_region   7891..8182
                     /note="AluSx repeat: matches 293. .2 of consensus"
     repeat_region   8554..8873
                     /note="MLT2CA repeat: matches 3. .338 of consensus"
     repeat_region   8810..8924
                     /note="MLT2CB repeat: matches 316. .434 of consensus"
     repeat_region   8924..9108
                     /note="MSTA repeat: matches 1. .193 of consensus"
     repeat_region   9111..9366
                     /note="AluSx repeat: matches 42. .300 of consensus;
                     incomplete repeat"
     repeat_region   9383..9562
                     /note="MSTA repeat: matches 243. .426 of consensus"
     repeat_region   9562..9646
                     /note="MLT2G repeat: matches 305. .387 of consensus"
     variation       9643..9645
                     /note="clone 92N15; ACA in this entry; substitution"
                     /replace="ata"
     variation       9686..9688
                     /note="clone 92N15; ATG in this entry; substitution"
                     /replace="acg"
     variation       9712..9714
                     /note="clone 92N15; GAA in this entry; substitution"
                     /replace="gta"
     repeat_region   9715..10013
                     /note="AluJo repeat: matches 300. .1 of consensus"
     repeat_region   11174..11260
                     /note="MIR2 repeat: matches 146. .59 of consensus"
     repeat_region   11282..11532
                     /note="MIR repeat: matches 1. .262 of consensus"
     repeat_region   11635..11945
                     /note="AluSx repeat: matches 302. .2 of consensus"
     repeat_region   12189..12353
                     /note="AluSg repeat: matches 150. .300 of consensus;
                     incomplete repeat"
     repeat_region   12599..12901
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   13045..13304
                     /note="AluJb repeat: matches 297. .39 of consensus;
                     incomplete repeat"
     repeat_region   13608..13891
                     /note="AluJo repeat: matches 1. .300 of consensus"
     repeat_region   14005..14306
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   15239..15273
                     /note="MER3 repeat: matches 200. .166 of consensus"
     repeat_region   15274..15452
                     /note="AluJo repeat: matches 123. .301 of consensus;
                     incomplete repeat"
     repeat_region   15453..15583
                     /note="MER3 repeat: matches 162. .32 of consensus"
     repeat_region   15586..15867
                     /note="AluJb repeat: matches 5. .291 of consensus"
     repeat_region   15989..16142
                     /note="LIMD1 repeat: matches 961. .807 of consensus"
     repeat_region   16145..16443
                     /note="LIMD2 repeat: matches 875. .582 of consensus"
     repeat_region   16463..16765
                     /note="AluSx repeat: matches 302. .1 of consensus"
     repeat_region   17433..17734
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   17751..18071
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   18806..19089
                     /note="AluJo repeat: matches 292. .1 of consensus"
     repeat_region   19431..19732
                     /note="AluY repeat: matches 301. .1 of consensus"
     repeat_region   19892..20007
                     /note="FLAM_C repeat: matches 116. .1 of consensus"
     repeat_region   20197..20309
```

```
     repeat_region   20690..20979
                      /note="MIR repeat: matches 144..43 of consensus"
     repeat_region   21716..22001
                      /note="AluSx repeat: matches 302..10 of consensus"
     repeat_region   22137..22350
                      /note="AluSq repeat: matches 1..303 of consensus"
                      /note="AluJb repeat: matches 85..302 of consensus;
                      incomplete repeat"
     repeat_region   23113..23418
                      /note="AluJb repeat: matches 302..1 of consensus"
     repeat_region   23420..23605
                      /note="AluJb repeat: matches 207..14 of consensus"
     repeat_region   23635..23704
                      /note="MIR repeat: matches 207..14 of consensus"
     repeat_region   23773..23838
                      /note="MIR2 repeat: matches 128..59 of consensus"
     repeat_region   24258..24553
                      /note="MIR2 repeat: matches 81..146 of consensus"
     repeat_region   24560..24867
                      /note="AluSq repeat: matches 1..301 of consensus"
     repeat_region   24896..25041
                      /note="AluYb8 repeat: matches 1..308 of consensus"
     repeat_region   25407..25708
                      /note="FLAM_A repeat: matches 133..1 of consensus"
     unsure          26138
                      /note="AluSq repeat: matches 303..2 of consensus"
                      /note="this base could be T"
     repeat_region   26292..26647
                      /note="Li1 repeat: matches 1943..1788 of consensus"
     repeat_region   26469..26770
                      /note="MIR2 repeat: matches 302..1 of consensus"
     repeat_region   26847..27145
                      /note="AluJo repeat: matches 302..1 of consensus"

alignment_scores:
        Quality:       38.00              Length:       10
        Ratio:        3.800               Gaps:          0
Percent Similarity: 100.000     Percent Identity: 60.000

alignment_block:
US-08-653-294-13 x HS204E5

Align seg 1/1 to: HS204E5 from: 1 to: 129969

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       |||:|||::|||::|||::|||:::||
71948 TACAGGGTAGCAGTCAGGATGGACCACAGG 71977

seq_name: gb_pr4:AC009239

seq_documentation_block:
LOCUS       AC009239   136371 bp   DNA              PRI    22-OCT-1999
DEFINITION  Homo sapiens clone NH0470K20, complete sequence.
ACCESSION   AC009239
VERSION     AC009239.3  GI:6094637
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136371)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 136371)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (06-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 136371)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (22-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
COMMENT     MO 63108, USA
            On Oct 22, 1999 this sequence version replaced gi:5732147.
FEATURES             Location/Qualifiers
     source          1..136371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0470K20"
BASE COUNT    39683 a  25507 c  25721 g  45560 t
ORIGIN

alignment_scores:
        Quality:       38.00              Length:        9
        Ratio:        4.222               Gaps:          0
Percent Similarity: 100.000     Percent Identity: 77.778

alignment_block:
US-08-653-294-13 x AC009239

Align seg 1/1 to: AC009239 from: 1 to: 136371

     1 TyrArgLeuAlaIleArgLeuAspGlu 9
       |||:|||||||||||||||||||
124410 TATAAAGTTGCAATTAGATTGGATGAA 124436

seq_name: gb_htg4:AC011830

seq_documentation_block:
LOCUS       AC011830   166832 bp   DNA              20-OCT-1999
DEFINITION  Homo sapiens chromosome 8 clone 94_A_14 HTG map 8, *** SEQUENCING IN
            PROGRESS ***, 17 unordered pieces.
ACCESSION   AC011830
VERSION     AC011830.2  GI:6087976
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166832)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 8, clone 94_A_14
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 166832)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkly,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Connelan,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 20, 1999 this sequence version replaced gi:6041961.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997).
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *                 1     845: contig of 845 bp in length
```

```
    TITLE      The sequence of Homo sapiens clone
    JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 187447)
  AUTHORS   Waterston,R.H.
    TITLE     Direct Submission
    JOURNAL   Submitted (11-SEP-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
    COMMENT   On Oct 4, 1999 this sequence version replaced gi:5870314.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *           1     26045:  contig of 26045 bp in length
            *       26046     26063:  gap of unknown length
            *       26064     72624:  contig of 46561 bp in length
            *       72625     72642:  gap of unknown length
            *       72643    187447:  contig of 114805 bp in length.
FEATURES             Location/Qualifiers
     source          1..187447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0209Il1"
BASE COUNT    56242 a  35522 c  36150 g  59497 t     36 others
ORIGIN

alignment_scores:
        Quality:       38.00           Length:        9
        Ratio:          4.222          Gaps:          0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-13 x AC010086/rev

Align seg 1/1 to reverse of AC010086   from: 1  to: 187447

        1 TyrArgLeuAlaIleArgLeuAspGlu 9
          |||::::::|||||||||||||||||
    69009 TATAAGTTGCATTAGATTGGATGAA 68983

seq_name: gb_htg4:AC008591

seq_documentation_block:
LOCUS       AC008591   211769 bp    DNA             HTG      31-OCT-1999
DEFINITION  Homo sapiens chromosome 5 clone CIT-HSPC_575N7, *** SEQUENCING IN
            PROGRESS ***,  60 unordered pieces.
ACCESSION   AC008591
VERSION     AC008591.2  GI:6165161
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211769)
  AUTHORS   DOE Joint Genome Institute.
    TITLE     Sequencing of Human Chromosome 5
    JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211769)
  AUTHORS   DOE Joint Genome Institute.
    TITLE     Direct Submission
    JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
    COMMENT   On Oct 31, 1999 this sequence version replaced gi:5586476.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 60 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
*           846         gap of unknown length
*          2883:   contig of 2038 bp in length
*          2884         gap of unknown length
*          5685:   contig of 2802 bp in length
*          5686         gap of unknown length
*          9996:   contig of 4311 bp in length
*          9997         gap of unknown length
*         14822:   contig of 4826 bp in length
*         14823         gap of unknown length
*         19370:   contig of 4548 bp in length
*         19371         gap of unknown length
*         24594:   contig of 5224 bp in length
*         24595         gap of unknown length
*         30146:   contig of 5552 bp in length
*         30147         gap of unknown length
*         35542:   contig of 5396 bp in length
*         35543         gap of unknown length
*         44699:   contig of 9157 bp in length
*         44700         gap of unknown length
*         55420:   contig of 10721 bp in length
*         55421         gap of unknown length
*         67431:   contig of 12011 bp in length
*         67432         gap of unknown length
*         84061:   contig of 16630 bp in length
*         84062         gap of unknown length
*        100053:   contig of 15992 bp in length
*        100054         gap of unknown length
*        123949:   contig of 23896 bp in length
*        123950         gap of unknown length
*        161245:   contig of 37296 bp in length
*        161246         gap of unknown length
*        166832:   contig of 5587 bp in length.
FEATURES             Location/Qualifiers
     source          1..166832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="94_A_14"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT   45111 a  39419 c  39077 g  42526 t   699 others
ORIGIN

alignment_scores:
        Quality:       38.00           Length:        9
        Ratio:          4.222          Gaps:          0
Percent Similarity: 100.000   Percent Identity: 77.778

alignment_block:
US-08-653-294-13 x AC011830

Align seg 1/1 to AC011830 from: 1  to: 166832

        1 TyrArgLeuAlaIleArgLeuAspGlu 9
          |||||||||::|||||:::::|||||
    91099 TACAGGCTGAGCATCCGAGTGGATGAG 91125

seq_name: gb_htg3:AC010086

seq_documentation_block:
LOCUS       AC010086   187447 bp   DNA              HTG      04-OCT-1999
DEFINITION  Homo sapiens clone NH0209Il1, *** SEQUENCING IN PROGRESS ***, 3
            unordered pieces.
ACCESSION   AC010086
VERSION     AC010086.2  GI:6007895
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187447)
  AUTHORS   Waterston,R.H.
```

```
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*        1    518: contig of 518 bp in length          34614   35537: contig of 924 bp in length
*      519   1379: gap of unknown length               35538   37384: gap of unknown length
*     1380   1992: contig of 613 bp in length          37385   38975: contig of 1591 bp in length
*     1993   3019: gap of unknown length               38976   42433: gap of unknown length
*     3020   4116: contig of 1097 bp in length         42434   46186: contig of 3753 bp in length
*     4117   4468: gap of unknown length               46187   48955: gap of unknown length
*     4469   4869: contig of 401 bp in length          48956   52157: contig of 3202 bp in length
*     4870   6986: gap of unknown length               52158   57534: gap of unknown length
*     6987   8350: contig of 1364 bp in length         57535   60590: contig of 3056 bp in length
*     8351   9169: gap of unknown length               60591   66089: gap of unknown length
*     9170  10576: contig of 1407 bp in length         66090   70445: contig of 4356 bp in length
*    10577  10743: gap of unknown length               70446   77265: gap of unknown length
*    10744  11269: contig of 526 bp in length          77266   82716: contig of 5451 bp in length
*    11270  11921: gap of unknown length               82717   87967: gap of unknown length
*    11922  12533: contig of 612 bp in length          87968   95151: contig of 7184 bp in length
*    12534  12829: gap of unknown length               95152  102796: gap of unknown length
*    12830  13906: contig of 1077 bp in length        102797  111023: contig of 8227 bp in length
*    13907  14105: gap of unknown length              111024  119350: gap of unknown length
*    14106  14914: contig of 809 bp in length         119351  129014: contig of 9664 bp in length
*    14915  16489: gap of unknown length              129015  140858: gap of unknown length
*    16490  16732: contig of 243 bp in length         140859  154309: contig of 13451 bp in length
*    16733  16933: gap of unknown length              154310  164838: gap of unknown length
*    16934  17739: contig of 806 bp in length         164839  177853: contig of 13015 bp in length
*    17740  17875: gap of unknown length              177854  192072: gap of unknown length
*    17876  19318: contig of 1443 bp in length        192073  211769: contig of 19697 bp in length.
*    19319  21247: gap of unknown length
*    21248  22282: contig of 1035 bp in length
*    22283  23232: gap of unknown length
*    23233  24221: contig of 989 bp in length
*    24222  25716: gap of unknown length
*    25717  27600: contig of 1884 bp in length
*    27601  28647: gap of unknown length
*    28648  30401: contig of 1754 bp in length
*    30402  31662: gap of unknown length
*    31663  34613: contig of 2951 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..211769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CIT-HSPC_575N7"
BASE COUNT       64860 a    39430 c    40173 g    66628 t       678 others
ORIGIN
```

```
alignment_scores:
     Quality:   38.00                    Length:      10
     Ratio:      4.222                    Gaps:         0
     Percent Similarity:  90.000         Percent Identity:  70.000

alignment_block:
US-08-653-294-13 x AC008591/rev

Align seg 1/1  to reverse of: AC008591  from: 1  to: 211769

        1 TyrArgLeuAlaIleArgLeuAspGluArg 10
          ||||||| :::||||||||||:::|||
    30417 TACCGGCTTATAGTTCGGCTCGACTCTAGA 30388
```

seq_name: gb_inl:AF026152

```
seq_documentation_block:
LOCUS       AF026152      486 bp    DNA              INV       26-OCT-1997
DEFINITION  Caenorhabditis elegans chitinase (CHT1) gene, partial cds.
ACCESSION   AF026152
VERSION     AF026152.1  GI:2564714
KEYWORDS    .
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 486)
AUTHORS     de la Vega,H., Specht,C.A., Liu,Y. and Robbins,P.W.
TITLE       Chitinases are a multi-gene family in Aedes, Anopheles, and
            Drosophila
JOURNAL     Insect Mol. Biol. (1997) In press
REFERENCE   2  (bases 1 to 486)
AUTHORS     Specht,C.A. and de la Vega,H.
TITLE       Direct Submission
JOURNAL     Submitted (19-SEP-1997) Biology, Massachusetts Institute of
            Technology, 40 Ames St. E17-235, Cambridge, MA 02139, USA
FEATURES    Location/Qualifiers
     source 1..486
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
     mRNA   join(<1..14,150..>486)
            /gene="CHT1"
            /product="chitinase"
     gene   <1..>486
            /gene="CHT1"
     CDS    join(<1..14,150..>486)
            /gene="CHT1"
            /codon_start=1
            /product="chitinase"
            /protein_id="AAB81847.1"
            /db_xref="GI:2564715"
            /translation="WAQYRQGRAKFVPEDYTPGLCTHILFAFGWMNADYTVRAYDPAD
            LPNDWAGEGMYRRVNLKVTDQLKTLLSFGGCSFGTALFQGMAASSASRKVFIDSAT
            TSVRTWGFDGIDLDW"
BASE COUNT  130 a   90 c  120 g  146 t
ORIGIN

alignment_scores:
        Quality:     37.00          Length:      10
        Ratio:        4.111         Gaps:         0
Percent Similarity:  90.000   Percent Identity:  70.000

alignment_block:
US-08-653-294-13 x AF026152         ..

Align seg 1/1  to: AF026152   from: 1  to: 486

         1 TyrArgLeuAlaIleArgLeuaspGluArg 10
           ||| |||:||||| :::|||||||||
       204 TATTCTCTTTGCATTCGGATGATGAACGC 233
```

seq_name: gb_bal:HSMOXMXU

```
seq_documentation_block:
LOCUS       HSMOXMXU      1323 bp   DNA              BCT       10-APR-1997
DEFINITION  Hyphomicrobium sp. DNA, Mox mutant mxu-1::Tn5-132.
ACCESSION   Y09104
VERSION     Y09104.1  GI:1869806
KEYWORDS    methanol oxidation system; Mox.
SOURCE      Hyphomicrobium sp.
ORGANISM    Hyphomicrobium sp.
            Bacteria; Proteobacteria; alpha subdivision; Hyphomicrobium.
REFERENCE   1  (bases 1 to 1323)
AUTHORS     Gliesche,C.G., Menzel,M. and Fesefeldt,M.
TITLE       A rapid method for creating species-specific gene probes for
            methylotrophic bacteria
JOURNAL     J. Microbiol. Methods 28, 25-34 (1997)
REFERENCE   2  (bases 1 to 1323)
AUTHORS     Gliesche,C.G.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-1996) C.G. Gliesche, CAU-Kiel, Institut fuer
            Allgemeine Mikrobiologie, Am Botanischen Garten 1-9,
            (Biologiezentrum), D-24118 Kiel, FRG
FEATURES    Location/Qualifiers
     source 1..1323
            /organism="Hyphomicrobium sp."
            /strain="B 69"
            /db_xref="taxon:82"
     mRNA   <1..>1323
            /note="mxu-1::Tn5-132; fragment of methanol oxidation
            system (Mox)"
BASE COUNT  221 a  402 c  429 g  271 t
ORIGIN

alignment_scores:
        Quality:     37.00          Length:       8
        Ratio:        4.625         Gaps:         0
Percent Similarity: 100.000   Percent Identity:  87.500

alignment_block:
US-08-653-294-13 x HSMOXMXU         ..

Align seg 1/1  to: HSMOXMXU   from: 1  to: 1323

         1 TyrArgLeuAlaIleArgLeuAsp  8
           ||||||||||| :::|||||||||
       967 TATCGCCTGGCGCCTCGTCTCGAT 990
```

seq_name: gb_inl:CELC04F6

```
seq_documentation_block:
LOCUS       CELC04F6      25083 bp  DNA              INV       19-DEC-1995
DEFINITION  Caenorhabditis elegans cosmid C04F6.
ACCESSION   U42835
VERSION     U42835.1  GI:1125760
KEYWORDS    .
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans strain=Bristol N2.
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 25083)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
REFERENCE   2  (bases 1 to 25083)
AUTHORS     Nhan.M.
TITLE       The sequence of C. elegans cosmid C04F6
JOURNAL     Unpublished (1995)
REFERENCE   3  (bases 1 to 25083)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (10-DEC-1995) Robert Waterston
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
```

```
                    Sanger Centre, Hinton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                    NEIGHBORING COSMID INFORMATION:
        The 5' cosmid is ZK563, 200 bp overlap;3' cosmid is ZK813, 200 bp
        overlap. Actual start of this cosmid is at base position 197 of
        CELCO4F6; actual end is at 20314 of CELZK813
        NOTES:
        Coding sequences below are predicted from computer analysis, using
        the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES          Location/Qualifiers
     source       1..25083
                  /organism="Caenorhabditis elegans"
                  /db_xref="taxon:6239"
                  /map="X"
     gene          357..673
                  /gene="C04F6.2"
     CDS           join(357..422,476..673)
                  /gene="C04F6.2"
                  /codon_start=1
                  /evidence=not_experimental
                  /protein_id="AAA83585.1"
                  /db_xref="GI:1125761"
                  /translation="MNTVYVGVTYKWLPETAGKTPQEINRHFSPEFPGTNVFLYLKHK
                  ASKTFAYLSGKPALMNFISFVVQLACCIFLLDFGVNIYKNIGK"
     gene          complement(1004..3221)
                  /gene="C04F6.3"
     CDS           complement(join(1004..1347,1404..2716,2852..2966,
                  3140..3221))
                  /gene="C04F6.3"
                  /note="coded for by C. elegans cDNA CEESK79R; coded for by
                  C. elegans cDNA yk104c7.3; coded for by C. elegans cDNA
                  yk109d2.3; coded for by C. elegans cDNA yk106b4.3; coded
                  for by C. elegans cDNA yk84f3.3; coded for by C. elegans
                  cDNA yk51h2.3; coded for by C. elegans cDNA yk106b7.3;
                  coded for by C. elegans cDNA yk77f7.3; coded for by C.
                  elegans cDNA yk60g3.3; coded for by C. elegans cDNA
                  yk9d7.3; coded for by C. elegans cDNA CEESK79FB; coded for
                  by C. elegans cDNA yk9d7.3; coded for by C. elegans cDNA
                  yk9d7.5; coded for by C. elegans cDNA yk60g3.5; coded for
                  by C. elegans cDNA yk104h2.5; coded for by C. elegans cDNA
                  yk109d2.5; coded for by C. elegans cDNA yk84f3.5; coded
                  for by C. elegans cDNA yk51h2.5; coded for by C. elegans
                  cDNA yk104c7.5; coded for by C. elegans cDNA CEESS83F;
                  Similar to glycosyl hydrolase (family 18)."
                  /codon_start=1
                  /protein_id="AAA83586.1"
                  /db_xref="GI:1125762"
                  /translation="MLLGKFLLVASFIIPIAYTWTGATIRNHPADVVAARNKITSRSV
                  ARSEPTNSYIRPCYFTNWAQYRQGRAKFVPEDYTPGLCTHILFAFGWMNADYTVRAYD
                  PADLPNDWAGEGMYRRVVNKLKVTDTQLKTLLSFGGWSFGTALFQGMAASSASRKVFID
                  SATFVRTWGFDGIDIDWEYPSCATDMANYVALVKELKAACESEAGSTGKDRLLVTAA
                  VAAGPAAIDAGYDIPNLAPNFDFILLMSYDFFGAWASLVGFNSPLYATTELPAEWNGW
                  NVDSSARYWNQKGMPKEKIIVGMPTYGRGWTLNNASAINPGTSGGSPAKITQYVQEAGV
                  GAYFEFCMLANGATRYWDSQSOVPYIVQGNOWWSYDDEESFANKMAYVKREGYGGAF
                  VWTLDFDDFNAGCSNSNGQLYPLISVIAKELGGVIIPKKGGVTTAPTTVATTVTTGRP
                  PMTSAVTTTNAATTTRAATTTASNTNVCSGKSDGFYPNSNNCGLFVLCLSSKSYS
                  MSCPSGLQYSASLKYCTTSTASCCSVTTTRAPTTTKSAPTVTTTRAPTTTTPAFKC
                  TKDGFFGVPSDCLKFIRCVNGISYNFECPNGLSFHADTMMCDRPDPSKCAK"
     gene          6686..11760
                  /gene="vit-5"
     CDS           join(6686..7549,7620..7904,7952..10989,11039..11267,
                  11313..11446,11501..11760)
                  /gene="vit-5"
                  /note="coded for by C. elegans cDNA yk31e4.5; coded for by
                  C. elegans cDNA CEMSF35F; coded for by C. elegans cDNA
                  CEMSG37FB; coded for by C. elegans cDNA CEMSH27R; coded
                  for by C. elegans cDNA CEMSE22F; coded for by C. elegans
                  cDNA yk32f10.5; coded for by C. elegans cDNA CEMSG03F;
                  coded for by C. elegans cDNA CEMSG79F; coded for by C.
                  elegans cDNA CEMSH88F; coded for by C. elegans cDNA
                  CEMSF29F; coded for by C. elegans cDNA CEMSF29F; coded
```

```
                  for by C. elegans cDNA CEMSF59R; coded for by C. elegans
                  cDNA CEMSH88R; coded for by C. elegans cDNA CEMSH78R;
                  coded for by C. elegans cDNA yk1e5.5; coded for by C.
                  elegans cDNA cm08f5; coded for by C. elegans cDNA
                  yk16c5.5; coded for by C. elegans cDNA yk19b10.5; coded
                  for by C. elegans cDNA yk42f1.5; coded for by C. elegans
                  cDNA yk41f9.5; coded for by C. elegans cDNA yk16c5.3;
                  coded for by C. elegans cDNA yk32f10.3; coded for by C.
                  elegans cDNA yk19b10.3; coded for by C. elegans cDNA
                  yk45e8.5; coded for by C. elegans cDNA yk31h2.3; coded for
                  by C. elegans cDNA CEMSB21F; coded for by C. elegans cDNA
                  yk41f9.3; coded for by C. elegans cDNA yk19e1.5; C04F6.1."
                  /codon_start=1
                  /product="C. elegans vitellogenin 5 precursor, vit-5
                  (Spieth et al., NAR 13:7129-7138(1985))"
                  /protein_id="AAA83587.1"
                  /db_xref="GI:1125763"
                  /translation="MKSIIIASLVALAIAASPALDRIFSPKSEYVYKFDGLLLSGLPT
                  ASSDASQTLISCRTRLQAVDDRYIHLQLTDIQYSASHIPQSEQWPKIESLEQRELSDE
                  FKELELLPFRAQIRNGLISEIQFSSEDAEWSKNAKRSILNLFSLRKSAPVDEMNQDQK
                  DMESDKDSLFFNVHEKTMEGDCEVAYIVQEGEKTIITKSVNFDKCITRPETAYGLRF
                  GSECKECEKEGQFVKPQTVYTYTFKNERLQESEHVSVTLNVNGQEVVKSETRAKVTF
                  VEESKINREIKKVSGPKEIVYSMENEKLIEQYQQGKAENVPFKAIEMEQKVEQLQ
                  EIFRQIOEHEQNTPETVHLIAARVRMFRMCTIEELKKVHTTIYTKAEKKVQLVIETTL
                  AVAGTKNTIQHLIHHFEKKSITPLRAAELLKSVQETLYSEHIADLLIQLAQSPLSEK
                  YEPLRQSAWLAAGSVVRGFASKTQDLPLIPNSEWLSGVQFDFATIFERNAFLP
                  LALKTLGNAGIDLSVYELVQIIQDPROPLSIRTEAVDALRLLKDVMPRKIQKVLLPVY
                  KNRQNKPELRMAALWRMMHTIPEEPVLAHIVSQMENESNQHVAAFTVNVLRQFYKSTN
                  PCTQQLAVRCSKILLFTRYQPQEQMLSTYSQLPFNSEWLSGVQFDFATIFEKNAFLP
                  KEVQASEPTVFGGNWNKYFAQVGFSQONFEQVILKTLEKLSLYGKQSDELRSRRVQSG
                  IQMLQEIVKKMNIRPRVQQTDSQNAHAVFLRYKEMDYIVLPIDMEIIDTLVEKYVRN
                  GEPDIKSLLTFLINDSKFELHRALFFYEAERRIPTTIGMPLTISGKMPTILSINGKVS
                  IELEKLGARLVLDIVPTVATTHVTEMRFWYPVIEQGVKSLQSARLHTPLRFESTVELK
                  KNTLEITHKFVVPENKKTTVSVHTRPVAFIRVPKNQDSEYVEAEEKISHSQYQMSTE
                  EIDRQYETFGLRINAQGNVLSOWTLPMVLMTEQDFEYTLENKNRPVEFTARVTIGNLE
                  KTDLSEIKFDKIFEKEFDLENNESENRRQYFHKMIREIQSEOGFKNLITLKLEAPOQM
                  YWNTELRTVCDWIRMCKVEMDARRSPMEHENKEWTLRTELLAARPOMPSSLRQLREQ
                  PHREVQLAFNAWGSSKKSEITVNAQLEQSTEQKFIRNIEREYKGIPEYELLIKAAR
                  LNQVNVVSEYKILTPQSEYTFSRIFDLIKAYNFWTVSEKRVQNENRRVVLQLSVEPLSR
                  QYMNMTIQTPEQVELKNVRIPRVVLFPTIARSAMFOQTWEKTGATCKVDQSEVSTFDN
                  VIYRAPLTTCYSLVAKDCSEQPRFAVLAKKINKNSEELLVKVRREEIVKKSDDKF
                  LVKVDGKKVNPTELEQYNIEILGDNLIVIRLPQGEVRFDGYTVKTNMPSVASQNQLCG
                  LCGNNDGERDNEMTADNYETEDVEEFHRSYILKNEECEVENDRISEKKNYRNKWNRE
                  EKKSDYESSSDYSNYDEKETEKELVKTLIKFSNRVCFSIEPVSECRRGLESEKTS
                  NKKIRFTCMPRHSKNARRFLKPAREQTVAELVDFPVSFVESVKIPTACVAY"
     gene          complement(12024..14819)
                  /gene="C04F6.4"
     CDS           complement(join(12024..12130,12178..12364,12413..12909,
                  12952..13444,13873..14351,14747..14819))
                  /gene="C04F6.4"
                  /note="coded for by C. elegans cDNA yk21f1.5; coded for by
                  C. elegans cDNA yk21f1.3; coded for by C. elegans cDNA
                  yk54a8.3; coded for by C. elegans cDNA yk45e8.5; Similar
                  to beta transducin."
                  /codon_start=1
                  /protein_id="AAA83588.1"
                  /db_xref="GI:1125764"
                  /translation="MSEFSQTALFPSLPRTARGTAVVLGNTPAGDKIQYCNGTSVYTV
                  PVGSLTDEIYTEHSHQTVAKTSPSGYCASGDVHGNVRIWDTQTTHILKTTIPVF
                  SGPVKDISWDSESKRIAVGEGRERFGHVFLPDTGTSNGNLTGQARAMNSVDFKPSRP
                  FRIISGSDDNTVAIFEGPPFKFKSTFGEHTKFVHSVRYNPDGSLFASTGGDGTIVLYN
                  GVDGTKTGVFEDDSLKNVAHSGSVFGLFWSPDGTKIASASADKTIKIWNVATLKVEKT
                  IPVGTRIEDQQLGIIWTKQALVSISANGFINFVNPELGSIDQVRYGHNKAITALSSSA
                  DGKTLFSADAEGHINSWDISTGISNRVFPDVHATMITGIHKTTSKGLFTVSWDDHLKV
                  VPAGGSGVDSSKAVANKLSSQPLGLAVSADGDIAVAACYKHIAYSHGKLTEVPISYN
                  SCVALSNDKQFVAVGGQDSKVHVYKLSGASVSEVKTIVHPAEITVAFSNNGAFLVA
                  TQDSRKVIPYSVANNFELAHTNSWTFHTAKVACVSWSPDNVRLATGSLDNSVIVWNMN
                  KPSDHPIIIKGAHAMSSVNSVIWLNETTIVSAQDSNIKFWNVPF"
     gene          complement(16182..22204)
                  /gene="C04F6.5"
     CDS           complement(join(16182..16307,16354..16515,16970..17083,
                  17154..17225,17270..17373,17416..17649,17758..17911,
                  20223..20626,20679..21012,21062..21256,21312..21497,
```

```
                21547..21613,21669..21772,22130..22204))
                /gene="C04F6.5"
                /note="Similar to alcohol/ribitol dehydrogenase."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA83589.1"
                /db_xref="GI:1125765"
                /translation="MESKVVALNRESRRWKKSSFNELLKRARNQKSTICNRFTRKNTP
                NILVFTSRLTLRPFFNFACWHSAMEIFESMSFSVKFDKELLPIGRHGLTAGWLIE
                SLGQLLNSDNIRNIKCSQLGTDGFVSQIMRVVIEFTNNQIKKYIVKMPETNIKAALE
                KTTQQKMPEGADEQFIGGLTMFFNREVEYYAMEKIPGLRTPKCYHAQQWDNGKTTGAI
                VMQDLEGMVSVPYYESLNLQQIASIGSQMLNLHLFSIGMSDEWRQKFPFMELIDTVS
                NMTDIYKLYVSRNPELEEGFSKVEKMYQDRELFTKVLRDSHKSLGIDDFLCHGDLWFY
                NLMWIPRKSGSEVASNHLGSIIDWQNVHTGNICEDFCHMLTFCCDTEIRRQAENTFLP
                YYYNQLKAKAVEAGKKLNLTLNQFLRAYRRNFIAHALHLPFIVSIMLCVKPADDEIVQ
                KIRNQMSGALMSAAQFAVTSYVCVRVLKFLYIMCKSVLVHFITPKHDLDYLKDTWTVI
                TGGTDGIGKAYIEELCKTRGLKKFYLIGRNIDKLNNTKKELVEQHGCEVMCVHDFEK

alignment_scores:
        Quality:       37.00              Length:        10
        Ratio:         4.111              Gaps:          0
Percent Similarity:    90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x CELC04F6/rev   ..

Align seg 1/1  to reverse of: CELC04F6  from: 1  to: 25083

    1 TyrArgLeuAlaIleArgAspGluArg 10
      ||| |||::||||||:|||||||||
 2662 TATTCTCTTGCATTCGGATGGATGAACGC 2633

seq_name: gb_htg5:AC015346

seq_documentation_block:
LOCUS       AC015346   26121 bp   DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015346
VERSION     AC015346.1  GI:6435989
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 26121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210169 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..26121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     6182 a   6662 c   6750 g   6507 t
ORIGIN

alignment_scores:
        Quality:       37.00              Length:         9
        Ratio:         4.625              Gaps:          0
Percent Similarity:    88.889   Percent Identity: 77.778

alignment_block:
US-08-653-294-13 x AC015346   ..
```

```
Align seg 1/1  to: AC015346  from: 1  to: 26121

    1 TyrArgLeuAlaIleArgLeuAspGlu 9
      ||||||:::|||  ||||||||||||
14973 TATCGGATCGCTACGCGTCTTGACGAA 14999
```

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-13 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000   1:27 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV-xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-08-653-294-13
Query length: 10
Database: N_Geneseq_36:*
Database sequences: 111585
Database length: 125096042
Search time (sec): 590.520000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:T80415	-	37.00	104.57	317.99	13987	! Hybrid srmG/tylG ORF1. DNA en
N_Geneseq_36:T78508	-	37.00	94.16	1.2e+03	44377	! Platenolide synthase gene clu
N_Geneseq_36:T80414	-	37.00	94.16	1.2e+03	44377	! Platenolide synthase gene clu
N_Geneseq_36:T04154	-	34.00	111.17	136.43	1791	! BCG DapB gene. Gene involved i
N_Geneseq_36:T06770	+	34.00	102.66	406.40	4602	! Pseudomonas aureofaciens phena
N_Geneseq_36:T89957	+	34.00	102.65	406.50	4603	! Pseudomonas aureofaciens phena
N_Geneseq_36:V58732	+	34.00	102.65	406.50	4603	! Phenazine gene cluster. Genes
N_Geneseq_36:V39844	+	34.00	100.73	520.30	5698	! Pseudomonas fluorescens phenaz
N_Geneseq_36:X51895	+	33.00	123.99	26.35	278	! Human secreted protein 5' EST
N_Geneseq_36:V74662	+	33.00	107.06	230.95	1816	! Staphylococcus aureus contig S
N_Geneseq_36:T15599	+	33.00	102.90	393.69	2880	! Laccase-LCC2 gene. DNA constru
N_Geneseq_36:V83943	+	33.00	90.18	2.0e+03	11811	! Bacterial artificial chromosc
N_Geneseq_36:T07040	+	32.00	112.89	109.35	612	! Immunogen DNA from n-(ABCDE) he
N_Geneseq_36:T34189	+	32.00	108.72	186.72	972	! PcpD coding sequence. Flavobact
N_Geneseq_36:V84117	+	32.00	108.13	201.46	1038	! Pseudomonas aeruginosa heptos
N_Geneseq_36:T09298	+	32.00	107.42	220.44	1122	! Bovine p32 11-cis-retinol dehy
N_Geneseq_36:T68935	+	32.00	107.42	220.44	1122	! Bovine p32 11-cis-retinol dehy
N_Geneseq_36:X35241	+	32.00	107.42	220.44	1122	! cDNA encoding p32 protein whic
N_Geneseq_36:T84457	+	32.00	103.16	380.81	1800	! KOD-1 heat shock protein gene.
N_Geneseq_36:X00596	+	32.00	103.14	382.04	1805	! Thermococcus sp. KS-1 strain H
N_Geneseq_36:Q55147	+	32.00	93.51	1.3e+03	5247	! Pseudomonas aeruginosa probe F
N_Geneseq_36:T84508	+	32.00	90.96	1.8e+03	6961	! Rabbit LDL receptor analogue c
N_Geneseq_36:X06749	+	32.00	84.65	4.1e+03	14024	! E. coli O157 antigen gene clu
N_Geneseq_36:X13037	+	31.50	75.00	1.4e+04	32768	! Enterococcus faecalis genome
N_Geneseq_36:Q95197	+	31.00	115.31	80.17	301	! Simple tandem repeat (STR) corr
N_Geneseq_36:V20113	+	31.00	114.32	91.04	336	! Probe (14) for microbial genes
N_Geneseq_36:T88153	+	31.00	104.88	305.51	957	! S. marcesens Ssp autotransporte
N_Geneseq_36:X06864	+	31.00	101.97	443.94	1322	! Australian banana cv. Williams
N_Geneseq_36:Q20995	+	31.00	100.47	537.62	1560	! EHV-4 gC gene. Nucleic acid se
N_Geneseq_36:V35473	+	31.00	96.36	911.33	2462	! Human hSK3 coding sequence. DN
N_Geneseq_36:Q25594	+	31.00	96.27	921.19	2485	! Rabbit GnT I cDNA clone. Human
N_Geneseq_36:X08414	+	31.00	96.15	936.64	2521	! Human small conductance calciu
N_Geneseq_36:V49983	+	31.00	95.74	986.23	2636	! Floral organ-specific chitinas
N_Geneseq_36:X20531	+	31.00	92.31	1.5e+03	3858	! Polynucleotide sequence from t
N_Geneseq_36:V65326	+	31.00	92.04	1.6e+03	3973	! Human neuronatin genomic DNA s
N_Geneseq_36:N81546	+	31.00	91.16	1.8e+03	4382	! Bio A, Bio B and Bio D-encodin
N_Geneseq_36:V58938	+	31.00	83.63	4.7e+03	10095	! Mycobacterium tuberculosis em
N_Geneseq_36:V74378	+	31.00	78.24	9.7e+03	18355	! Staphylococcus aureus contig
N_Geneseq_36:Q40706	+	31.00	73.48	1.7e+04	31122	! Bacillus subtilis srfA operon
N_Geneseq_36:V22674	+	31.00	69.14	2.9e+04	50341	! DNA sequence of a L5 shuttle
N_Geneseq_36:T51411	+	31.00	68.80	3.0e+04	52297	! Mycobacteriophage L5 genome s

N_Geneseq_36:Q47357	-	31.00	68.80	3.0e+04	52298	! L5 mycobacteriophage DNA.
N_Geneseq_36:V30458_0	-	31.00	62.09	7.0e+04	110000	! Rhizobium species plasm1
N_Geneseq_36:X20248_00	-	31.00	62.09	7.0e+04	110000	! Borrelia burgdorferi po
N_Geneseq_36:V30459_0	-	31.00	62.09	7.0e+04	110000	! Rhizobium species symbio

seq_name: N_Geneseq_36:T80415

seq_documentation_block:
ID T80415 standard; DNA; 13987 BP.
AC T80415;
DT 02-MAR-1998 (first entry)
DE Hybrid srmG/tylG ORF1.
KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW platenolide synthase gene cluster; platenolide production; srmG gene;
KW polyketide; tylactone synthesis; antibiotic; tylosin; hybrid gene; ss.
OS Streptomyces ambofaciens.
OS Streptomyces fradiae.
FH Key   Location/Qualifiers
FT CDS   350..13987
FT   /tag= a
FT   /transl_except= (pos:350..352, aa:Met)
FT   /note= "ORF1 encodes hybrid protein shown in W22611"
PN EP-791655-A2.
PD 27-AUG-1997.
PF 19-FEB-1997; 301056.
PR 22-FEB-1996; US-012078.
PA (ELIL) LILLY & CO ELI.
PI Dehoff BS, Kuhstoss SA,   Rosteck PR,   Sutton KL;
DR WPI; 97-418046/39.
DR P-PSDB: W22611.
PT DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT production of tylosin-related polyketide compounds
PS Claim 22; Pages 178-197; 220pp; English.
CC This sequence represents a hybrid gene of the invention. This sequence
CC was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a
CC EcoRI-SfuI fragment from tylG ORF1. The position of the nucleotides from
CC each of the two genes is not given in the specification. The srmG gene
CC (see T80414) was isolated from Streptomyces ambofaciens, and encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The tylG gene (see T80413) is the tylactone
CC synthase gene cluster of the invention. The tylG sequence was isolated
CC from Streptomyces fradiae, and encodes multifunctional proteins which
CC direct the synthesis of the polyketide tylactone. Tylactone is the basic
CC building block of the antibiotic tylosin. The hybrid sequence can be used
CC to transform S. ambofaciens lacking the srmG ORF1 sequence, or S. fradiae
CC lacking the tylG ORF1 sequence, so that they can produce polyketides. The
CC DNA sequence can be modified so as to alter the type of carboxylic acids
CC incorporated, the number of carboxylic acids incorporated and/or the
CC post-condensation reactions performed, thereby resulting in novel
CC tylosin-related polyketides.
SQ Sequence 13987 BP;   1556 A;   4401 C;   5727 G;   2303 T;

alignment_scores:
                   Quality: 37.00   Length: 9
                   Ratio: 4.625   Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 88.889

alignment_block:
US-08-653-294-13 x T80415/rev

Align seg 1/1 to reverse of: T80415 from: 1 to: 13987

        1 TyrArgLeuAlaIleArgLeuAspGlu 9
          ||||||||||||| |||||||||||||||
     6652 TACCGCCTCGCCCAGCGCCTGACGAG 6626

seq_name: N_Geneseq_36:T78508

seq_documentation_block:
ID T78508 standard; DNA; 44377 BP.
AC T78508;

```
DT   26-FEB-1998 (first entry)
DE   Platenolide synthase gene cluster.
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
OS   Streptomyces ambofaciens.
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in W23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in W23717"
FT   CDS             20010..31284
FT                   /*tag= c
FT                   /note= "ORF3 encodes protein shown in W23718"
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in W23719"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in W23720"
PN   EP-791656-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301066.
PR   22-FEB-1996; US-012050.
PA   (ELIL) LILLY & CO ELI.
PI   Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI   Rosteck PR;
DR   WPI: 97-418047/39.
DR   P-PSDB; W23716-W23720.
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
PS   Claim 9; Pages 8-33; 81pp; English.
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
SQ   Sequence 44377 BP;  4965 A; 15552 C; 17381 G; 6479 T;

alignment_scores:
        Quality:        37.00           Length:          9
        Ratio:          4.625           Gaps:            0
Percent Similarity: 88.889     Percent Identity: 88.889

alignment_block:
US-08-653-294-13 x T78508/rev

Align seg 1/1  to reverse of: T78508  from: 1  to: 44377

    1 TyrArgLeuAlaIleArgLeuAspGlu 9
      ||||||||||  ||||||||||
 6667 TACCGCCCTCGCCCAGCGCCTGGACGAG 6641

seq_name: N_Geneseq_36:T80414

seq_documentation_block:
ID   T80414 standard; DNA; 44377 BP.
AC   T80414;
DT   27-FEB-1998 (first entry)
DE   Platenolide synthase gene cluster.
KW   Tylactone synthase gene cluster; tylG gene; platenolide production; srmG gene;
```

```
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
OS   Streptomyces ambofaciens.
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in W22606"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in W22607"
FT   CDS             20010..31284
FT                   /*tag= c
FT                   /note= "ORF3 encodes protein shown in W22608"
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in W22609"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in W22610"
PN   EP-791655-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301056.
PR   22-FEB-1996; US-012078.
PA   (ELIL) LILLY & CO ELI.
PI   Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
DR   WPI: 97-418046/39.
DR   P-PSDB; W22606-W22610.
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
PS   Example 2; Pages 110-134; 220pp; English.
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see T80413) to create a hybrid ORF1 sequence (see T80415). The tylG gene
CC   is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
SQ   Sequence 44377 BP;  4965 A; 15552 C; 17381 G; 6479 T;

alignment_scores:
        Quality:        37.00           Length:          9
        Ratio:          4.625           Gaps:            0
Percent Similarity: 88.889     Percent Identity: 88.889

alignment_block:
US-08-653-294-13 x T80414/rev

Align seg 1/1  to reverse of: T80414  from: 1  to: 44377

    1 TyrArgLeuAlaIleArgLeuAspGlu 9
      ||||||||||  ||||||||||
 6667 TACCGCCCTCGCCCAGCGCCTGGACGAG 6641

seq_name: N_Geneseq_36:T04154

seq_documentation_block:
ID   T04154 standard; DNA; 1791 BP.
AC   T04154;
DT   26-FEB-1996 (first entry)
DE   BCG DapB gene.
```

```
KW   Dihydrodipicolinate-reductase; diaminopimelate-dehydrogenase; dapB;
KW   diaminopimelic acid; peptidoglycan; mycobacteria; vaccine;
KW   Mycobacterium tuberculosis; Mycobacterium avium;
KW   Mycobacterium fortuitum; Mycobacterium leprae; Mycobacterium gordoneae;
KW   Mycobacterium haemophilium; Mycobacterium paratuberculosis; BCG; ss.
OS   Mycobacterium bovis
FH   Key             Location/Qualifiers
FT   rbs             297..303
FT                   /*tag= a
FT   cds             312..1127
FT                   /*tag= b
FT                   /product= DapB
FT   rbs             1136..1140
FT                   /*tag= c
FT   cds             1151..1683
FT                   /*tag= d
FT                   /product= unidentified
PN   WO9523226-Al.
PD   31-AUG-1995.
PF   27-FEB-1995; U02455.
PR   28-FEB-1994; US-203190.
PA   (YESH ) UNIV, YESHIVA EINSTEIN COLLEGE.
PI   Cirillo JD, Jacobs WR;
DR   P-PSDB; R79946.
     WPI: 95-311535/40.
PT   Gene involved in the synthesis of di:amino:pimelic acid in
PT   mycobacteria - and methods for inhibiting its expression to treat
PT   mycobacterial infection
PS   Claim 3; Fig 3; 48pp; English.
CC   A BCG DNA fragment (T04154) that complemented a dapB mutation in
CC   Escherichia coli was isolated.  The BGC dapB gene encoded a
CC   bifunctional enzyme (R79946) capable of catalyzing the
CC   dihydrodipicolinate-reductase and diaminopimelate-dehydrogenase
CC   reactions involved in diaminopimelic acid (DAP) biosynthesis. DAP
CC   biosynthetic genes are useful as targets for anti-mycobacterial
CC   agents and for the design of in vivo selection systems. Mycobacteria
CC   having a deleted dapB gene can be used for vaccine prodn.
SQ   Sequence  1791 BP;  254 A;  638 C;  640 G;  259 T;

alignment_scores:
       Quality:     34.00          Length:          9
       Ratio:        3.778          Gaps:            0
Percent Similarity: 100.000     Percent Identity: 66.667

alignment_block:
US-08-653-294-13 x T04154/rev    ..

Align seg 1/1 to reverse of: T04154    from: 1   to: 1791

     1 TyrArgLeuAlaIleArgAspGlu 9
       :::|||||||:::||||||||:::
   716 TTTCGTCTCGGATGCGTCTGACGAT 690

seq_name: N_Geneseq_36:T06770

seq_documentation_block:
ID   T06770 standard: DNA; 4602 BP.
AC   T06770;
DT   15-OCT-1996 (first entry)
DE   Pseudomonas aureofaciens phenazine gene cluster.
KW   Phenazine; biosynthesis; phenazine; antibiotic;
KW   Antipathogenic substance; ss.
KW   fungicide; pesticide; ss.
OS   Pseudomonas aureofaciens.
FH   Key             Location/Qualifiers
FT   cds             230..1597
FT                   /*tag= a
FT                   /note= "phz1 (ORF1)"
FT   cds             1598..2761
FT                   /*tag= b
FT                   /note= "phz2 (ORF2)"
FT   cds             2764..3600
FT                   /*tag= c
```

```
FT                   /note= "phz3 (ORF3)"
FT   cds             3597..4265
FT                   /*tag= d
FT                   /note= "phz4 (ORF4)"
PN   WO9533818-A2.
PD   14-DEC-1995.
PF   30-MAY-1995; IB0414.
PR   08-JUN-1994; US-258261.
PA   (CIBA ) CIBA GEIGY AG.
PI   Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;
PI   Ligon J, Ryals JA, Schupp T, Uknes SJ;
DR   WPI: 96-040226/04.
DR   P-PSDB; R87533, R87534, R87535, R87536.
PT   New genes for biosynthesis of anti-pathogenic substances - pref.
PT   pyrrolnitrin and soraphen, useful for disease control in plants
PS   Disclosure; Page 163-169; 190pp; English.
CC   This is the phenazine gene cluster encoding 4 open reading frames
CC   which each encode 1 polypeptide. The gene cluster may be expressed
CC   recombinantly to produce phenazine, or expressed in a transgenic
CC   plant for disease-resistance.
SQ   Sequence  4602 BP;  925 A;  1516 C;  1341 G;  818 T;

alignment_scores:
       Quality:     34.00          Length:         10
       Ratio:        3.778          Gaps:            0
Percent Similarity:  90.000     Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x T06770    ..

Align seg 1/1 to: T06770    from: 1   to: 4602

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||||||||||  :::||||:::||:||:::
   994 TATCGACTTGCCTGCCACCTCGATCGGCGC 1023

seq_name: N_Geneseq_36:T89957

seq_documentation_block:
ID   T89957 standard; DNA; 4603 BP.
AC   T89957;
DT   12-MAR-1998  (first entry)
DE   Pseudomonas aureofaciens phenazine gene cluster genomic DNA.
KW   Phenazine; biosynthesis; antibiotic; antipathogenic;
KW   transgenic plant; phytopathogen; resistance; ss.
OS   Pseudomonas aureofaciens.
FH   Key             Location/Qualifiers
FT   CDS             230..1597
FT                   /*tag= a
FT                   /product= phz1
FT   CDS             1598..2761
FT                   /*tag= b
FT                   /product= phz2
FT                   /transl_except= (pos: 1796..1798, aa: Xaa)
FT                   /note= "Xaa = unknown"
FT   CDS             2764..3600
FT                   /*tag= c
FT                   /product= phz3
FT   CDS             3597..4265
FT                   /*tag= d
FT                   /product= phz4
FT                   /transl_except= (pos: 3391..3393, aa: Xaa)
FT                   /note= "Xaa = unknown"
PN   US5662898-A.
PD   02-SEP-1997.
PF   01-JUN-1995; 457342.
PR   20-AUG-1990; US-570184.
PR   02-JUL-1992; US-908284.
PR   31-AUG-1992; US-937648.
PR   01-JUL-1993; US-087636.
PR   08-JUN-1994; US-258261.
PA   (CIBA ) CIBA GEIGY CORP.
```

```
PI   Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;
PI   Ligon JM, Ryals JA, Schupp T, Uknes SJ;
DR   WPI; 97-447901/41.
DR   P-PSDB; W31304; W31305; W31306; W31307.
PT   protecting plants against pathogens with genetically transformed
PT   biological control agent - which expresses all polypeptide(s)
PT   involved in pyrrolnitrin biosynthetic pathway
PS   Example 18; Column 135-144; 88pp; English.
CC   This genomic DNA sequence encodes a cluster of genes involved in
CC   phenazine biosynthesis. Phenazines are nitrogen-containing
CC   heterocyclic compounds with a common planar aromatic tricyclic
CC   structure. It has been proposed that phenazine antibiotic function
CC   arises from the formation of intercalative complexes with DNA
CC   interfering with DNA metabolism. Transgenic plants containing such
CC   antipathogenic genes should have enhanced resistance to attack by
CC   phytopathogens.
SQ   Sequence   4603 BP;   925 A;   1516 C;   1342 G;   818 T;

alignment_scores:
             Quality:    34.00          Length:      10
             Ratio:      3.778          Gaps:         0
Percent Similarity:     90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13  to: T89957  x T89957

Align seg 1/1  to: T89957  from: 1  to: 4603

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||||||||||  :::|||||::|||
   994 TATCGACTTGCCTGCCACCTCGATCGGCGC 1023

seq_name: N_Geneseq_36:V58732

seq_documentation_block:
ID   V58732 standard; DNA; 4603 BP.
AC   V58732;
DT   07-DEC-1998 (first entry)
DE   Phenazine gene cluster.
KW   Pyrrolnitrin: biosynthetic pathway; pathogen protection; phenazine;
KW   plant antipathogenic substance production: anti-fungal antibiotic;
KW   fungal respiratory electron transport inhibitor; lipoprotein damage; ss.
OS   Pseudomonas aureofaciens.
FH   Key          Location/Qualifiers
FT   CDS          230..1597
FT                  /*tag= a
FT                  /product= phz1
FT   CDS          1598..2761
FT                  /*tag= b
FT                  /product= phz2
FT   CDS          2764..3600
FT                  /*tag= c
FT                  /product= phz3
FT   CDS          3597..4265
FT                  /*tag= d
FT                  /product= phz4
PN   US5817502-A.
PD   06-OCT-1998.
PF   09-OCT-1996; US-729214.
PF   09-OCT-1996; US-729214.
PR   09-OCT-1996; US-729214.
PR   08-JUN-1994; US-258261.
PR   30-MAY-1995; WO-IB0414.
PA   (NOVS ) NOVARTIS FINANCE CORP.
PI   Hammer PE, Hill DS, Kirner S, Lam ST, Ligon JM,
PI   Van Pee K;
WPI; 98-556391/47.
DR   P-PSDB; W69401, W69402, W69403, W69404.
PT   Genes encoding enzymes of the biosynthetic pathway of pyrrolnitrin -
PT   useful for producing transgenic plants which can produce
PT   pyrrolnitrin as an anti-pathogenic agent
PS   Example 18: Column 147-156; 109pp; English.
CC   This sequence represents the phenazine gene cluster, isolated from
```

```
CC   Pseudomonas aureofaciens. This sequence was used to isolate the
CC   pyrrolnitrin gene region of the invention, that encodes at least one
CC   enzyme required in the biosynthetic pathway of pyrrolnitrin. The DNA and
CC   host cells transformed with it are useful for the production of
CC   transgenic plants with protection against phytopathogens. The enzymes
CC   are part of a biosynthetic pathway producing plant antipathogenic
CC   substances (APS). The compound ultimately produced by the pathway,
CC   pyrrolnitrin, is a broad range, anti-fungal antibiotic. It inhibits
CC   fungal respiratory electron transport and causes general lipoprotein
CC   damage. The transformed cells can additionally be used in compositions to
CC   be applied to plants to provide resistance, as can purified APS produced
CC   by them. Transgene dependent resistance eliminates the need to spray
CC   crops with chemical based pesticides and antibiotics, which is expensive
CC   and time consuming, and in addition, especially in the case of
CC   antibiotics, their over use leads to resistance. In addition, transgenic
CC   production of these enzymes overcomes problems of applying
CC   micro-organisms which can be slow growing and isolated in their growth
CC   areas.
SQ   Sequence   4603 BP;   925 A;   1516 C;   1342 G;   818 T;

alignment_scores:
             Quality:    34.00          Length:      10
             Ratio:      3.778          Gaps:         0
Percent Similarity:     90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13  to: V58732  x V58732

Align seg 1/1  to: V58732  from: 1  to: 4603

     1 TyrArgLeuAlaIleArgLeuAspGluArg 10
       ||||||||||  :::|||||::|||
   994 TATCGACTTGCCTGCCACCTCGATCGGCGC 1023

seq_name: N_Geneseq_36:V39844

seq_documentation_block:
ID   V39844 standard; DNA; 5698 BP.
AC   V39844;
DT   29-SEP-1998 (first entry)
DE   Pseudomonas fluorescens phenazine gene cluster.
KW   Pseudomonas; genetic engineering; biocontrol; plant; pathogenic;
KW   Rhizoctonia; Pythium; antifungal; pyrrolnitrin; crop protection; ss.
OS   Pseudomonas fluorescens
FH   Key          Location/Qualifiers
FT   CDS          105..1307
FT                  /*tag= a
FT                  /product= "phzF"
FT                  /note= "ORF1"
FT   CDS          1323..1946
FT                  /*tag= b
FT                  /product= "phzA"
FT                  /note= "ORF2"
FT   CDS          1943..3856
FT                  /*tag= c
FT                  /product= "phzB"
FT                  /note= "ORF3"
FT   CDS          3859..4695
FT                  /*tag= d
FT                  /product= "phzC"
FT                  /note= "ORF4"
FT   CDS          4692..5360
FT                  /*tag= e
FT                  /product= "phzD"
FT                  /note= "ORF5"
PN   WO9824919-A1.
PD   11-JUN-1998.
PF   05-DEC-1997; E06815.
PR   05-DEC-1997; US-058304.
PR   09-SEP-1997; US-058304.
PR   06-DEC-1996; US-761258.
PA   (NOVS ) NOVARTIS AG.
PI   Gaffney TD, Hill DS, Lam ST, Ligon JM, Stafford JM,
```

PI Torkewitz NR;
DR WPI; 98-333337/29.
PT Genetically modified Pseudomonas strains - useful to protect crop
PT plants by controlling or inhibiting plant pathogen growth, e.g.
PT growth of Rhizoctonia species
PS Example 9; Page 71-75; 85pp; English.
CC A genetically engineered biocontrol strain of Pseudomonas has been
CC developed that can control attacks on crop plants by pathogenic fungi,
CC e.g. Rhizoctonia and Pythium and aggressively compete with indigenous
CC bacteria and microflora in the plant rhizosphere. The strains can be
CC included with agronomically acceptable carriers or chemical fungicides
CC (e.g. metalaxyl compounds) in biocontrol compositions. The strains or
CC compositions can be applied to a plant/plant part to protect it from a
CC plant pathogenic fungus, by controlling or inhibiting fungal growth.
CC They can also be applied to the environment in which a plant pathogenic
CC fungus will grow (e.g. soil) to similarly control or inhibit pathogen
CC growth, or to seeds to protect seeds developing from the seed from a
CC plant pathogenic fungus. They are especially effective against
CC Rhizoctonia and Pythium species which cause damping off in cotton.
CC Rhizoctonia also infects many other crop species (e.g. beans and wheat),
CC and no effective chemical fungicides are available. The present sequence
CC represents the Pseudomonas fluorescens phenazine gene cluster used
CC in an example from the present invention.
SQ Sequence 5698 BP; 1166 A; 1855 C; 1677 G; 1000 T;

alignment_scores:
    Quality: 34.00   Length: 10
    Ratio: 3.778   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 70.000

alignment_block:
US-08-653-294-13 x V39844

Align seg 1/1 to: V39844 from: 1 to: 5698

```
   1 TyrArgLeuAlaIleArgLeuAspGluArg 10
     ||||||||||  :::||||||||:::|||
2089 TATCGACTTGCCTGCCACCTCGATCGGCGC 2118
```

seq_name: N_Geneseq_36:X51895

seq_documentation_block:
ID X51895 standard; DNA; 278 BP.
AC X51895;
DT 22-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO: 109.
KW Human: secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation: haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic: anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906552-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1236.
PR 01-AUG-1997; US-905223.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153782/13.
DR P-PSDB: Y13095.
PT New isolated brain-derived nucleic acids - used to develop products
PT which may have cytokine, immune, regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
PS Claim 1; Page 256; 577pp; English.
PT New isolated brain-derived sequence tags (ESTs) for human
PT secreted proteins, and encode the proteins given in Y12987 to Y13319,
PT respectively. The proteins given represent the signal peptide and an
PT N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation

CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 278 BP; 95 A; 58 G; 56 C; 67 T;

alignment_scores:
    Quality: 33.00   Length: 9
    Ratio: 4.125   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-08-653-294-13 x X51895

Align seg 1/1 to: X51895 from: 1 to: 278

```
  1 TyrArgLeuAlaIleArgLeuAspGlu 9
    ||||||||||:::||||:||:::|||
241 TACAGACTAGCTCGCGCATCACAGAG 267
```

seq_name: N_Geneseq_36:V74662

seq_documentation_block:
ID V74662 standard; DNA; 1816 BP.
AC V74662;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #351.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT misc_feature   241..300
FT   /*tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1230-1231; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences

CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 1816 BP;   539 A;   307 C;   240 G;   667 T;

alignment_scores:
        Quality: 33.00   Length: 9
          Ratio: 4.714   Gaps: 0
Percent Similarity: 77.778   Percent Identity: 77.778

alignment_block:
US-08-653-294-13 x V74662/rev   ..

Align seg 1/1 to reverse of: V74662   from: 1   to: 1816

       1 TyrArgLeuAlaIleArgAspGlu 9
         |||||||| ||| |||||||||
    1387 TATCGTTTACTAATAACATTAGACGAA 1361

seq_name: N_Geneseq_36:T15599

seq_documentation_block:
ID   T15599 standard; DNA; 2880 BP.
AC   T15599;
DT   07-APR-1996 (first entry)
DE   Laccase-LCC2 gene.
KW   Laccase-LCC2; Polyporus pinsitus; Trametes villosa; primer;
KW   polymerase chain reaction; PCR; signal peptide; cellulase;
KW   Aspergillus oryzae; cDNA probe; Escherichia coli; plasmid pDSY19;
KW   lignin; lignosulphonate; polymerisation; Kraft pulp;
KW   depolymerisation; oxidation; hair dye; phenol; aniline; vector;
KW   cloning; basidiomycetes; ss.
OS   Polyporus pinsitus.
OS   Trametes villosa.
FH   Key   Location/Qualifiers
FT   cds   364..2492
FT   /*tag= a
FT   /product= Laccase-LCC2
FT   /note= "EC-1.10.3.2"
FT   signal_peptide   364..423
FT   /*tag= b
FT   exon   364..543
FT   /*tag= c
FT   intron   544..592
FT   /*tag= d
FT   exon   593..661
FT   /*tag= e
FT   intron   662..715
FT   /*tag= f
FT   exon   716..835
FT   /*tag= g
FT   intron   836..899
FT   /*tag= h
FT   exon   900..1013
FT   /*tag= i
FT   intron   1014..1066
FT   /*tag= j
FT   exon   1067..1132
FT   /*tag= k
FT   intron   1133..1187
FT   /*tag= l
FT   exon   1188..1283
FT   /*tag= m
FT   intron   1284..1343
FT   /*tag= n
FT   exon   1344..1498
FT   /*tag= o
FT   intron   1499..1553
FT   /*tag= p
FT   exon   1554..1751
FT   /*tag= q
FT   intron   1752..1815

FT   exon   1816..1872
FT   /*tag= r
FT   intron   1873..1928
FT   /*tag= s
FT   exon   1929..2135
FT   /*tag= t
FT   intron   2136..2195
FT   /*tag= u
FT   exon   2196..2492
FT   /*tag= v
FT   /*tag= w
PN   WO9600290-A1.
PD   04-JAN-1996.
PF   15-JUN-1995; U07536.
PR   24-JUN-1994; US-265534.
PR   15-MAY-1995; US-441147.
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO-NORDISK AS.
PI   Aaslyng DA, Dalboge H, Schneider P, Xu F, Yaver DS;
DR   WPI; 96-068874/07.
DR   P-PSDB; R90722.
PT   DNA constructs for expression of Polyporus laccase enzymes - for use
PT   in e.g. lignin manipulation, juice mfr., phenol polymerisation and
PT   phenol resin prodn.
PS   Claim 5; Page 62-65; 137pp; English.
CC   The sequence encodes laccase-LCC2 (pI 5.95) from Polyporus pinsitus
CC   (Trametes villosa). Polymerase chain reaction (PCR) amplification of
CC   P. pinsitus cDNA using primers 3331 (T15603) and 3332 (T15604) gives
CC   a 1500 bp fragment, which is joined to a signal peptide sequence
CC   from a 43-kDa cellulase using primer pHD433 (T15605) and a pUC
CC   forward primer in PCR. Clones are expressed in Aspergillus oryzae,
CC   and a cDNA probe is obtained and used to screen a P. pinsitus
CC   genomic library in Escherichia coli DH5-alpha, giving plasmid
CC   pDSY19 (23GEN), with a 4-kb HindIII insert (NRRL B-21266).
CC   Screening also results in isolation of LCC1 (T15598) and LCC3-LCC5
CC   (T15600-T15602), which encode different laccases produced by P.
CC   pinsutus. The laccases may be used to polymerise lignin or
CC   lignosulphonates, to depolymerise Kraft pulp, to oxidise dyes or
CC   precursors, in hair dye compositions, or to polymerise or oxidise a
CC   phenolic or aniline compound. These new laccases are well-expressed
CC   in Aspergillus spp. (with vector integration in the genome), in
CC   contrast to previous basidiomycete laccases, which give low yields
CC   of recombinant enzyme.
SQ   Sequence 2880 BP;   547 A;   908 C;   793 G;   632 T;

alignment_scores:
        Quality: 33.00   Length: 9
          Ratio: 4.125   Gaps: 0
Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-08-653-294-13 x T15599/rev   ..

Align seg 1/1 to reverse of: T15599   from: 1   to: 2880

       2 ArgLeuAlaIleArgLeuAspGluArg 10
         || |::::::||||||||||||||||
    1733 CGGGTGCAAGTTCGTCTCGATGAGCGG 1707

seq_name: N_Geneseq_36:V83943

seq_documentation_block:
ID   V83943 standard; DNA; 11811 BP.
AC   V83943;
DT   03-MAR-1999 (first entry)
DE   Bacterial artificial chromosome (BAC)-F2 contig 3.
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; BAC; transgenic;
KW   human artificial chromosome; bacterial artificial chromosome; ss.
OS   Synthetic.
PN   WO9851790-A1.

PD 19-NOV-1998.
PF 13-MAY-1998; AU0352.
PR 26-AUG-1997; AU-008791.
PR 13-MAY-1997; AU-006784.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Cancilla MR, Choo K, Du Sart D;
DR WPI; 99-009773/01.
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
PS Claim 10; Page 195-203; 540pp; English.
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
SQ Sequence 11811 BP; 3014 A; 2459 C; 2433 G; 3905 T;

alignment_scores:
          Quality: 33.00   Length: 10
            Ratio: 3.667   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 50.000

alignment_block:
US-08-653-294-13 x V83943/rev

Align seg 1/1 to reverse of: V83943 from: 1 to: 11811

       1 TyrArgLeuAlaIleArgLeuAspGluArg 10
         ||||||:::::::::||| ||||||:::
    4851 TACAGAATTCAGTCAGGAGGATGAAAA 4822

seq_name: N_Geneseq_36:T07040

seq_documentation_block:
ID T07040 standard; DNA; 612 BP.
AC T07040;
DT 03-JUL-1996 (first entry)
DE Immunogen DNA from n-(ABCDE) hepatitis virus.
KW Non-A, non-B, non-C, non-D, non-E hepatitis virus; n-(ABCDE);
KW immunogen; antibody; vaccine; phage library; ds.
OS Non-A, non-B, non-C, non-D, non-E hepatitis virus JFA clone 4B11.
PN WO9532290-A2.
PD 30-NOV-1995.
PF 17-MAY-1995; U05980.
PR 20-MAY-1994; US-246986.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Kim JP, Purcell RH;
DR WPI; 96-020585/02.
PT New non-A, -B, -C, -D and -E (n-(ABCDE)) hepatitis DNA libraries -
PT used to develop prods. for the detection, diagnosis, prevention and
PT treatment of n-(ABCDE) hepatitis.
PS Disclosure; Page 113; 165pp; English.
CC The sequence represents clone 4B11 which encodes an immunogenic
CC polypeptide associated with non-A, non-B, non-C, non-D, non-E
CC (n-(ABCDE)) hepatitis virus infection, and is obtained by
CC preparing a phage from a phage lambda gt11 library of JFA DNA
CC (ATCC 75271) n-(ABCDE) hepatitis serum, plating to form plaques,
CC and screening the phage plaques for the production of polypeptides

CC immunoreactive with n-(ABCDE) serum. Inserted sequences in gt11
CC are expressed as beta-galactosidase fusion proteins. Clone 4B11
CC has 3 multiple insert sequences; each insert can be fractionated
CC into discrete sequences by restriction enzymes, or portions of the
CC inserts can be PCR amplified by sequence specific primers. Each
CC resulting individual region can be subcloned and immunoscreened.
CC This allows identification of specific regions responsible for
CC immunoreactivity. n-(ABCDE) hepatitis virus polypeptides can be
CC used for the production or detection of antibodies, and in
CC vaccines. The antibodies can be used for detection, diagnosis and
CC in passive immunotherapy. The DNA can be used in detection and
CC diagnosis, and as hybridisation probes for identification of
CC further n-(ABCDE) hepatitis virus coding sequences. Culture systems
CC producing the n-(ABCDE) polypeptides can be used in screening
CC studies.
SQ Sequence 612 BP; 153 A; 171 C; 161 G; 127 T;

alignment_scores:
          Quality: 32.00   Length: 9
            Ratio: 3.556   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
US-08-653-294-13 x T07040/rev

Align seg 1/1 to reverse of: T07040 from: 1 to: 612

       2 ArgLeuAlaIleArgLeuAspGluArg 10
         |||||||||:::::::|||||||:::|||
      49 CGCCTGGCCGTCGAACTCGATGACCGC 23

seq_name: N_Geneseq_36:T34189

seq_documentation_block:
ID T34189 standard; DNA; 972 BP.
AC T34189;
DT 25-SEP-1996 (first entry)
DE PcpD coding sequence.
KW Pentachlorophenol breakdown pathway; PcpC; PcpA; PcpB; environment;
KW tetrachloro-p-hydroquinone reductase; PCP-degrading enzyme complex;
KW Flavobacterium; PCP; 2,3,5,6-tetrachloro-p-hydroquinone; TeCH;
KW glutathione; 2,3,6-trichloro-p-hydroquinone; TrCH; food chain;
KW 2,6-dichloro-p-hydroquinone; wood preserving industry; fungicide;
KW pesticide; herbicide; disinfectant; ds.
OS Flavobacterium sp. Strain ATCC 39723.
FH Key   Location/Qualifiers
FT misc_difference 754
FT   /*tag=   a
FT   /note= "Residue not given in the specification, included
FT   to maintain open reading frame and to encode the
FT   amino acid given in the specification at this
FT   position (Glu)"
PN US5512478-A.
PD 30-APR-1996.
PF 23-MAR-1996; 856015.
PR 23-MAR-1992; US-856015.
PR 13-JUL-1992; US-914282.
PR 18-JUL-1994; US-276887.
PA (IDAH-) IDAHO RES FOUND INC.
PI Lange CC, Orser CS, Xun L;
DR WPI; 96:229872/23.
DR P-PSDB; R99487.
PT Flavobacterium sp. penta:chloro:phenol breakdown pathway genes and
PT enzymes - useful in the bio-remediation and dechlorination of PCP
PT contg. matter
PS Disclosure; Columns 61-64; 52pp; English.
CC The sequences given in T34189-90 encode proteins from the
CC pentachlorophenol (Pcp) breakdown pathway, designated PcpD and
CC PcpR. The PcpD gene was found to code for a 323 amino acid
CC polypeptide, mol. wt. 35942 daltons. Based on multiple sequence
CC alignments, PcpD belongs to the family of class 1 dioxygenase electron
CC transport proteins. PcpD was predicted to have three structural

```
CC    domains, one involved in binding flavin mononucleotide (FMN), the second
CC    responsible for binding reduced NADPH and the third forming a plant-
CC    ferredoxin-type (2Fe-2S) centre.  PcpD is classified as an iron
CC    sulphur flavoprotein-oxidoreductase, and has been designated the PCP
CC    4-monooxygenase reductase.  It is transcribed as a dicistronic message
CC    with PcpB (see also T34184).  The pcpR open reading frame encodes a 303
CC    amino acid protein with a mol. wt. of 33549 daltons, which possesses a
CC    helix-turn-helix motif in its N-terminal portion.  PcpR is thought to
CC    activate the transcription of pcpBD and pcpA.  In combination with pcpA,
CC    pcpB and pcpC, these enzymes catalyse the initial steps of PCP breakdown.
CC    These enzymes can specifically be used in the breakdown of PCP
CC    containing matter which persists in the environment and in food chains
CC    after its use in the wood preserving industry as a fungicide and
CC    pesticide, and in products such as herbicides and disinfectants.
SQ    Sequence    972 BP;    181 A;    310 C;    313 G;    168 T;
```

```
alignment_scores:
            Quality:   32.00              Length:        8
              Ratio:    4.000              Gaps:          0
Percent Similarity: 100.000    Percent Identity:  75.000

alignment_block:
US-08-653-294-13 x T34189/rev      ..

Align seg 1/1  to reverse of: T34189  from: 1  to: 972

        3 LeuAlaIleArgLeuAspGluArg 10
          |||||::::|||::::|||||||||
      192 CTCGCCCTGCGGGTTGACGAGCGA 169
```

```
seq_name: N_Geneseq_36:V84117

seq_documentation_block:
ID  V84117  standard; DNA; 1038 BP.
AC  V84117;
DT  15-MAR-1999  (first entry)
DE  Pseudomonas aeruginosa heptosyl transferase II waaF gene.
KW  WaaF gene; rfaF gene; lipopolysaccharide; heptosyl transferase II waaF gene.
KW  diagnosis; vaccine; heptosyl transferase II; ss.
OS  Pseudomonas aeruginosa strain PAO1.
PN  WO9850557-A1.
PD  12-NOV-1998.
PF  01-MAY-1998; CA0395.
PR  09-MAY-1997; US-046149.
PR  02-MAY-1997; US-045418.
PA  (UYGU-) UNIV GUELPH.
PI  Burrows LL, De Kievit TR, Lam JS, Matewish M, Walsh A;
DR  WPI; 99-034725/03.
DR  P-PSDB; W88211.
PT  Isolated P. aeruginosa waaC, waaG, waaF and waaP gene cluster -
PT  useful in the diagnosis or treatment of P. aeruginosa infections
PS  Claim 2; Fig 6; 6ipp; English.
CC  This is the nucleotide sequence of the waaF gene (rfaF gene) of
CC  the waa gene cluster of Pseudomonas aeruginosa PAO1.  It codes for
CC  WaaF (see W82211), a heptosyl transferase II that adds the second
CC  heptose residue onto the core oligosaccharide in the biosynthesis
CC  of the lipopolysaccharide inner core.  The four waa genes
CC  of P. aeruginosa (see V84116-19) are arranged contiguously in an
CC  operon with the gene order waaF, waaC, waaG and waaP.  The
CC  functions of the encoded proteins (see W82210-13) were tested by
CC  complementation of specific Salmonella typhimurium mutants, and
CC  knockout mutations of the genes in P. aeruginosa.  The waa nucleic
CC  acids or proteins can be used to diagnose a bacterial, especially a
CC  P. aeruginosa, infection in an animal.  They can further be used to
CC  screen for compounds that affect core lipopolysaccharide biosynthesis
CC  or assembly.  A claimed method of treating or preventing a bacterial
CC  infection comprises administering an agent that inhibits a waa gene
CC  to an animal.  A claimed vaccine for treating a bacterial infection
CC  includes one of WaaP, WaaC, WaaF and WaaG.
SQ  Sequence   1038 BP;    148 A;    348 C;    364 G;    178 T;
```

```
alignment_scores:
            Quality:   32.00              Length:        8
              Ratio:    4.000              Gaps:          0
Percent Similarity: 100.000    Percent Identity:  75.000

alignment_block:
US-08-653-294-13 x V84117      ..

Align seg 1/1  to: V84117  from: 1  to: 1038

        3 LeuAlaIleArgLeuAspGluArg 10
          |||||::::|||||::::|||||||
      630 CTGGCTGTTCGGCTCGAAGAACGA 653
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:45 ; Search time 209.03 Seconds
                                              (without alignments)
                                              3.980 Million cell updates/sec

Title: US-08-653-294-18
Perfect score: 58
Sequence: 1 YRLAIRRILRY 12

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SPTREMBL_12:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	37	63.8	259	11	Q60503	Q60503 cricetulus
2	37	63.8	321	5	O45795	O45795 caenorhabdi
3	36	62.1	141	2	O84514	O84514 chlamydia t
4	36	62.1	243	2	O85853	O85853 sphingomona
5	36	62.1	309	2	O86347	O86347 mycobacteri
6	36	62.1	725	5	O16382	O16382 caenorhabdi
7	36	62.1	795	5	Q22063	Q22063 caenorhabdi
8	36	62.1	1847	5	P91495	P91495 caenorhabdi
9	36	62.1	1899	10	Q9XEG1	Q9xeg1 gossypium h
10	35	60.3	142	2	Q92789	Q9z7s9 chlamydia p
11	35	60.3	325	2	Q9XB05	Q9xb05 myxococcus
12	35	60.3	350	5	O45803	O45803 caenorhabdi
13	35	60.3	683	5	Q9Y136	Q9y136 drosophila
14	35	60.3	1283	5	Q24393	Q24393 drosophila
15	35	60.3	1605	5	O96446	O96446 vairimorpha
16	34	58.6	61	12	Q9WHV4	Q9whv4 squash yell
17	34	58.6	111	2	P70780	P70780 anabaena sp
18	34	58.6	152	2	Q9WYM7	Q9wym7 thermotoga
19	34	58.6	261	2	Q9Z671	Q9z671 zymomonas m
20	34	58.6	280	2	O35043	O35043 bacillus su

21	34	58.6	285	5	P91068	P91068 caenorhabdi
22	34	58.6	377	4	O60704	O60704 homo sapien
23	34	58.6	735	5	O76912	O76912 drosophila
24	34	58.6	777	4	O43162	O43162 homo sapien
25	34	58.6	814	10	O64487	O64487 arabidopsis
26	34	58.6	1712	5	O96160	O96160 plasmodium
27	33	56.9	35	5	O61237	O61237 onchocerca
28	33	56.9	59	7	O78094	O78094 homo sapien
29	33	56.9	133	2	Q92903	Q92903 chlamydia p
30	33	56.9	168	1	O28893	O28893 archaeoglob
31	33	56.9	184	3	O13954	O13954 schizosacch
32	33	56.9	194	5	Q21212	Q21212 caenorhabdi
33	33	56.9	260	12	Q9YW71	Q9yw71 melanoplus
34	33	56.9	313	2	P97213	P97213 clostridium
35	33	56.9	326	2	Q9ZJ72	Q9zj72 helicobacte
36	33	56.9	334	5	O52631	O52631 clostridium
37	33	56.9	346	10	O24513	O24513 arabidopsis
38	33	56.9	352	2	Q9X816	Q9x816 streptomyce
39	33	56.9	427	5	Q22310	Q22310 caenorhabdi
40	33	56.9	452	2	Q59958	Q59958 streptococc
41	33	56.9	453	2	O83072	O83072 treponema p
42	33	56.9	458	5	Q19489	Q19489 caenorhabdi
43	33	56.9	532	5	P72007	P72007 mycobacteri
44	33	56.9	576	2	O84579	O84579 chlamydia t
45	33	56.9	650	2	O52818	O52818 amycolatops

ALIGNMENTS

RESULT   1
Q60503
ID   Q60503   PRELIMINARY;   PRT;   259 AA.
AC   Q60503;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE   P-GLYCOPROTEIN (FRAGMENT).
GN   PGP3.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 89261726.
RA   NG W.F., SARANGI F., ZASTAWNY R.L., VEINOT-DREBOT L., LING V.;
RT   "Identification of members of the P-glycoprotein multigene family.";
RL   Mol. Cell. Biol. 9:1224-1232(1989).
DR   EMBL; M25792; AAA53439.1; -.
DR   EMBL; M25789; AAA53439.1; JOINED.
DR   EMBL; M25790; AAA53439.1; JOINED.
DR   EMBL; M25791; AAA53439.1; JOINED.
DR   PFAM; PF00005; ABC_tran; 1.
FT   NON_TER   1   1
SQ   SEQUENCE   259 AA; 28543 MW; 021E9AF3 CRC32;

Query Match   63.8%; Score 37; DB 11; Length 259;
Best Local Similarity   80.0%; Pred. No. 18;
Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   2 RLAIRRILLR 11
             |||||| |:|
Db   162 RLAIRRALIR 171

RESULT   2
O45795
ID   O45795   PRELIMINARY;   PRT;   321 AA.
AC   O45795;
DT   01-JUN-1998 (TREMBLrel. 06, Created)
DT   01-JUN-1998 (TREMBLrel. 06, Last sequence update)

```
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T19C9.4 PROTEIN.
GN   T19C9.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z92972; CAB07489.1; -.
DR   PFAM; PF01604; 7tm_5; 1.
SQ   SEQUENCE 321 AA; 36504 MW; 70031B52 CRC32;
```

```
Query Match
Best Local Similarity   63.8%;          Score 37; DB 5; Length 321;
                        88.9%;          Pred. No. 22;
Matches   8;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy      1 YRLAIRRIL 9
          || |||||||
Db    298 YRAAIRRIL 306
```

```
RESULT 3
O84514
ID   O84514      PRELIMINARY;        PRT;     141 AA.
AC   O84514;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L17.
GN   RL17.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL; AE001323; AAC68107.1; -.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
DR   PFAM; PF01196; Ribosomal_L17; 1.
KW   Ribosomal protein.
SQ   SEQUENCE 141 AA; 16152 MW; 2570FFF7 CRC32;
```

```
Query Match
Best Local Similarity   62.1%;          Score 36; DB 2; Length 141;
                        54.5%;          Pred. No. 15;
Matches   6;  Conservative  4;  Mismatches  ...;  Indels  1;  Gaps  0;

Qy      2 RLAIRRILRY 12
          ||| |:::||
Db     64 RLAVGRLMVRY 74
```

```
RESULT 4
O85853
ID   O85853      PRELIMINARY;        PRT;     243 AA.
AC   O85853;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 26.5 KD PROTEIN.
OS   Sphingomonas aromaticivorans.
OG   Plasmid pNL1.
OC   Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
OC   Sphingomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F199;
RA   ROMINE M.F., STILLWELL L.C., WONG K.-K., THURSTON S.J., SISK E.C.,
RA   SENSEN C.W., GAASTERLAND T., SAFER J.D., FREDRICKSON J.K.;
RT   "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas
RT   aromaticivorans strain F199.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF079317; AAD03866.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE 243 AA; 26455 MW; 40CDFBF4 CRC32;
```

```
Query Match
Best Local Similarity   62.1%;          Score 36; DB 2; Length 243;
                        63.6%;          Pred. No. 26;
Matches   7;  Conservative  2;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      2 RLAIRRILRY 12
          ||| ||:::||
Db      5 RLAQRRVTIRY 15
```

```
RESULT 5
O86347
ID   O86347      PRELIMINARY;        PRT;     309 AA.
AC   O86347;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 33.5 KD PROTEIN.
GN   RV2776C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY III C.E., TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   PARKHILL J.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL008967; CAA15591.1; -.
DR   HSSP; P33164; 2PIA.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PFAM; PF00111; fer2; 1.
DR   PFAM; PF00175; oxidored_fad; 1.
KW   Hypothetical protein; Iron-sulfur.
SQ   SEQUENCE 309 AA; 33517 MW; B152B590 CRC32;

Query Match             62.1%;  Score 36;  DB 2;  Length 309;
Best Local Similarity   87.5%;  Pred. No. 33;
Matches    7;  Conservative  1;  Mismatches   0;  Indels   0;  Gaps  0;

Qy      1 YRLAIRRI 8
            |:||||
Db     65 YRIAIRRI 72


RESULT 6
ID   O16382     PRELIMINARY;     PRT;   725 AA.
AC   O16382;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   K12B6.2 PROTEIN.
GN   K12B6.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   JONES K., WOHLDMANN P.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016442; AAB5917.1; -.
SQ   SEQUENCE 725 AA; 84926 MW; A4847D75 CRC32;

Query Match             62.1%;  Score 36;  DB 5;  Length 725;
Best Local Similarity   58.3%;  Pred. No. 77;
Matches    7;  Conservative  2;  Mismatches   3;  Indels   0;  Gaps  0;

Qy      1 YRLAIRRILRY 12
            :  |||:|:|
Db    594 FSFAIRRIVLNY 605


RESULT 7
ID   Q22063     PRELIMINARY;     PRT;   795 AA.
AC   Q22063; Q93489;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T01C3.10 PROTEIN.
GN   T01C3.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WILD A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z81061; CAB02939.1; -.
DR   EMBL; Z78413; CAB02939.1; JOINED.
DR   EMBL; Z78413; CAB01667.1; -.
DR   EMBL; Z81061; CAB01667.1; JOINED.
DR   HSSP; P19491; 1GR2.
DR   PFAM; PF00060; 11g_chan; 1.
SQ   SEQUENCE 795 AA; 89703 MW; DD722166 CRC32;

Query Match             62.1%;  Score 36;  DB 5;  Length 795;
Best Local Similarity   50.0%;  Pred. No. 85;
Matches    6;  Conservative  3;  Mismatches   3;  Indels   0;  Gaps  0;

Qy      1 YRLAIRRILRY 12
            ||:::||  ||
Db      6 YRTSLRRLATRY 17


RESULT 8
ID   P91495     PRELIMINARY;     PRT;   1847 AA.
AC   P91495;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SIMILARITY TO RAT INTEGRAL MEMBRANE GLYCOPROTEIN GP120 PRECURSOR.
GN   T23H2.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WAMSLEY P., BRADSHAW H.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80033; AAC48199.1; -.
SQ   SEQUENCE 1847 AA; 199637 MW; B0A25E0F CRC32;
```

Query Match   62.1%;   Score 36;   DB 5;   Length 1847;
Best Local Similarity 58.3%;   Pred. No. 1.9e+02;
Matches   7;   Conservative   2;   Mismatches   3;   Indels   0;   Gaps   0;

Qy   1 YRLAIRRILLRY 12
          ||| |:||| |
Db   19 YRLNVPRVLLPY 30

RESULT 9
Q9XEG1
ID   Q9XEG1   PRELIMINARY;   PRT;   1899 AA.
AC   Q9XEG1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE CALLOSE SYNTHASE CATALYTIC SUBUNIT.
GN   CFL1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   STRAIN=CV. TEXAS MARKER-1; TISSUE=PRIMARY-STAGE COTTON FIBER;
RA   CUI X., SHIN H., BROWN R.M.;
RT   "Cotton CFL1 gene shows homology to the yeast beta-1,3-glucan synthase
RT   subunit FKS1.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL: AF085717; AAD25952.1; -.
SQ   SEQUENCE   1899 AA;   218627 MW;   E695145F CRC32;

Query Match   62.1%;   Score 36;   DB 10;   Length 1899;
Best Local Similarity 63.6%;   Pred. No. 2e+02;
Matches   7;   Conservative   2;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   2 RLAIRRILLRY 12
          ||| || ||::
Db   554 RLAVSRIFLRF 564

RESULT 10
Q927S9
ID   Q927S9   PRELIMINARY;   PRT;   142 AA.
AC   Q927S9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L17.
GN   RL17.
OS   Chlamydia pneumoniae.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RA   KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA   GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT   "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL: AE001647; AAD18764.1; -.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   142 AA;   16400 MW;   4839EC84 CRC32;

Query Match   60.3%;   Score 35;   DB 2;   Length 142;
Best Local Similarity 54.5%;   Pred. No. 24;
Matches   6;   Conservative   4;   Mismatches   4;   Indels   1;   Gaps   0;

Qy   2 RLAIRRILLRY 12
          :|| |:::||

Db   64 RIAIGRLMVRY 74

RESULT 11
Q9XB05
ID   Q9XB05   PRELIMINARY;   PRT;   325 AA.
AC   Q9XB05;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MEMBRANE ASSOCIATED PROTEIN.
GN   TAD.
OS   Myxococcus xanthus.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ER-15;
RA   PAITAN Y., ORR E., RON E.Z., ROSENBERG E.;
RT   "Genetic and functional analysis of genes required for the post-
RT   modification of the polyketide antibiotic TA of Myxococcus xanthus.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ132503; CAB46503.1; -.
SQ   SEQUENCE   325 AA;   35985 MW;   4CC64E85 CRC32;

Query Match   60.3%;   Score 35;   DB 2;   Length 325;
Best Local Similarity 58.3%;   Pred. No. 54;
Matches   7;   Conservative   1;   Mismatches   4;   Indels   0;   Gaps   0;

Qy   1 YRLAIRRILLRY 12
          ||| :||: | |||
Db   202 YRLTVDRFPLRY 213

RESULT 12
O45803
ID   O45803   PRELIMINARY;   PRT;   350 AA.
AC   O45803;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   T23D5.1 PROTEIN.
GN   T23D5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LLOYD C., WILKINSON J.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; 282051; CAB04815.1; -.
DR   PFAM; PF01461; 7tm_4; 1.
SQ   SEQUENCE   350 AA;   40291 MW;   A9117B4D CRC32;

Query Match   60.3%;   Score 35;   DB 5;   Length 350;

Best Local Similarity   66.7%;   Pred. No. 58;
Matches   6; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

Qy   1 YRLAIRRIL 9
          ||||:|:|:|
Db   307 YRLAVRKIV 315

RESULT 13
Q9Y136
ID   Q9Y136;
AC   Q9Y136;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   BCDNA.GH07188.
GN   BCDNA.GH07188.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RUBIN G.M., WAN K.H., HARVEY D., LEWIS S.E., BROKSTEIN P., TSANG G.,
RA   AGBAYANI A., ARCANIA T.T., BAXTER E., BLAZEJ R.G., BUTENHOFF C.,
RA   CHAMPE M., CHAVEZ C., CHEW M., DOYLE C.M., FARFAN D.E., FRISE E.,
RA   GALLE R., GEORGE R.A., HARRIS N.L., HOSKINS R.A., EVANS-HOLM M.,
RA   HOUSTON K.A., HUMMASTI S.R., KIM E., LI P., MOSHREFI M., PACLEB J.M.,
RA   PARK S., SEQUEIRA A., SETHI H., SNIR E., SVIRSKAS R.R., WEINBURG T.,
RA   CELNIKER S.E.;
RT   "Full length Drosophila melanogaster cDNA sequence.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF145636; AAD38611.1; -.
SQ   SEQUENCE   683 AA; 79171 MW; 72C620E0 CRC32;

Query Match   60.3%; Score 35; DB 5; Length 683;
Best Local Similarity   66.7%; Pred. No. 1.1e+02;
Matches   8; Conservative   1; Mismatches   1; Indels   3; Gaps   0;

Qy   1 YRLAIRRILRY 12
          || :|||:|| ||
Db   206 YRNNIQRLQRY 217

RESULT 14
Q24393
ID   Q24393;
AC   Q24393;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   P-GLYCOPROTEIN/MULTIDRUG RESISTANCE PROTEIN.
GN   MDR50.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RX   MEDLINE; 94010914.
RA   GERRARD B., STEWART C., DEAN M.;
RT   "Analysis of Mdr50: a Drosophila P-glycoprotein/multidrug resistance
RT   gene homolog.";
RL   Genomics 17:83-88(1993).
DR   EMBL; L07065; AAA16186.1; -.
DR   FLYBASE; FBgn0010241; Mdr50.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1283 AA; 142538 MW; 65012909 CRC32;

Query Match   60.3%; Score 35; DB 5; Length 1283;
Best Local Similarity   70.0%; Pred. No. 2.1e+02;
Matches   7; Conservative   2; Mismatches   1; Indels   0; Gaps   0;

Qy   2 RLAIRRILLR 11
          |:||||| |:|
Db   547 RIAIRRALIR 556

RESULT 15
O96446
ID   O96446;   PRELIMINARY;   PRT;   1605 AA.
AC   O96446;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   LARGEST SUBUNIT OF RNA POLYMERASE II.
GN   RPB1.
OS   Vairimorpha necatrix.
OC   Eukaryota; Microsporidia; Burenellidae; Vairimorpha.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99110933.
RA   HIRT R.P., LOGSDON J.M. JR., HEALY B., DOREY M.W., DOOLITTLE W.F.,
RA   EMBLEY T.M.;
RT   "Microsporidia are related to Fungi: evidence from the largest subunit
RT   of RNA polymerase II and other proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:580-585(1999).
DR   EMBL; AF060234; AAD12604.1; -.
SQ   SEQUENCE   1605 AA; 180946 MW; 2D013184 CRC32;

Query Match   60.3%; Score 35; DB 5; Length 1605;
Best Local Similarity   58.3%; Pred. No. 2.6e+02;
Matches   7; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   1 YRLAIRRILRY 12
          ||:|:||| |
Db   1009 YNLSIKRILNEY 1020

Search completed: February 8, 2000, 13:17:46
Job time: 32495 sec

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-18 to: GenEmbl:*   out_format : pfs
Date: Feb 8, 2000 4:42 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15779/app-query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-18
Query length: 12
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 11370.480000

gb_in1:PPU6381			+	38.00	134.56	17.95	230	! AJ006381 Platythyrea punctat
gb_pl2:AF130141			-	38.00	126.81	48.54	541	! AF130141 Fusarium solani f.
gb_bal:UEU131275			-	38.00	122.55	83.80	865	! AJ131275 uncultured euryarch

seq_name: gb_htg7:AC013772

seq_documentation_block:
LOCUS   AC013772   142796 bp   DNA   HTG   09-DEC-1999
DEFINITION   Homo sapiens clone RP11-7O24, WORKING DRAFT SEQUENCE, 6 unordered
             pieces.
ACCESSION   AC013772
VERSION   AC013772.2   GI:6553994
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 142796)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-7O24
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 142796)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boguslavkly,L., Boukhgalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forres,C., Funke,R., Gage,D.,
             Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
             McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
             Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
             Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
             Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Dec 10, 1999 this sequence version replaced gi:6625750.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ----------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
             ----------- Project Information
                Center project name: L2959
                Center clone name: 7_O-24
             ----------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 128649 bases at least Q40
                Consensus quality: 137112 bases at least Q30
                Consensus quality: 140758 bases at least Q20
                Insert size: 138000; agarose-fp
                Insert size: 142796; sum-of-contigs
                Quality coverage: 5.7 in Q20 bases; agarose-fp
                Quality coverage: 5.5 in Q20 bases; sum-of-contigs
             ----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 6 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *   1   2142: contig of 2142 bp in length
             *   gap of unknown length

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_htg7:AC013772		47.00	110.79	378.51	142796	! AC013772 Homo sapiens clone
gb_htg2:HSDJ655C5	+	45.00	96.17	2.5e+03	306999	! AL121956 Homo sapiens chromo
gb_htg5:AC011145	-	44.00	98.46	1.8e+03	155243	! AC011145 Homo sapiens clone
gb_htg4:AC012150	+	42.00	103.73	936.58	37459	! AC012150 Homo sapiens chromos
gb_in1:DMRNA3	+	41.00	132.31	23.96	1050	! V00238 Three Drosophila melanc
gb_pr2:MMHLAFHOM	+	41.00	132.26	24.12	1056	! Z21819 M.mulatta HLA-F like mR
gb_htg7:AC017966	+	41.00	112.98	286.02	8849	! AC017966 Drosophila melanogast
gb_htg5:AC013927	+	41.00	109.65	438.30	12771	! AC013927 Drosophila melanogas
gb_bal:MLCB2533	+	41.00	99.24	1.7e+03	40245	! AL035310 Mycobacterium leprae
gb_bal:U00017	-	41.00	98.82	1.8e+03	42157	! U00017 Mycobacterium leprae c
gb_htg6:AC007807	+	41.00	88.56	6.5e+03	130536	! AC007807 Drosophila melanoga
gb_htg4:AC010564	-	41.00	87.68	7.3e+03	143914	! AC010564 Drosophila melanoga
gb_htg2:AC005047	+	41.00	86.67	8.1e+03	159103	! AC005047 Homo sapiens chromo
gb_pr4:AC007687	+	41.00	86.46	8.6e+03	164655	! AC007687 Homo sapiens 3q26.2
gb_htg1:HSA392M18	+	41.00	83.99	1.2e+04	216149	! AL121897 Homo sapiens chromo
gb_in2:AC005847	-	41.00	82.03	1.5e+04	268369	! AC005847 Drosophila melanoga
gb_htg3:AC008835	-	40.00	105.61	735.67	13052	! AC008835 Homo sapiens chromos
gb_in1:CEF14B6	-	40.00	98.28	1.9e+03	29283	! Z81502 Caenorhabditis elegans
gb_pl1:SPBC725	+	40.00	95.93	2.5e+03	37949	! AL034352 S.pombe chromosome I
gb_pl1:AB028605	-	40.00	95.90	2.6e+03	38089	! AB028605 Arabidopsis thaliana
gb_htg5:AC015514	+	40.00	89.30	6.0e+03	78785	! AC015514 Homo sapiens clone 2
gb_htg6:AC015746	-	40.00	87.95	7.1e+03	91513	! AC015746 Homo sapiens clone R
gb_htg5:AC015908	-	40.00	81.92	1.5e+04	177802	! AC015908 Homo sapiens chromc
gb_htg1:CEY47H10	+	40.00	77.28	2.8e+04	296589	! Z95311 Caenorhabditis elegan
gb_pl2:CNS019TC	-	39.00	136.00	14.94	300	! AL112072 Botrytis cinerea stral
gb_ro:MMU16741	+	39.00	120.31	111.62	1690	! U16741 Mus musculus capping pI
gb_ba2:AF026541	+	39.00	120.30	111.77	1692	! AF026541 Mycobacterium tubercu
gb_pr1:AB007931	-	39.00	107.23	597.75	7150	! AB007931 Homo sapiens mRNA for
gb_htg3:AC009605	+	39.00	92.57	3.9e+03	36000	! AC009605 Leishmania major chi
gb_in1:CELF45E1	+	39.00	92.05	4.2e+03	38122	! U28732 Caenorhabditis elegans
gb_bal:MTCY05A6	-	39.00	91.93	4.3e+03	38631	! Z96072 Mycobacterium tubercul
gb_htg5:AC012873	+	39.00	91.41	4.5e+03	40890	! AC012873 Drosophila melanogas
gb_htg5:AC013205	+	39.00	88.52	6.6e+03	56237	! AC013205 Drosophila melanogas
gb_pr3:AC005365	-	39.00	84.65	1.1e+04	86130	! AC005365 Homo sapiens chromos
gb_htg1:HS1141E20	-	39.00	83.49	1.3e+04	97906	! AL109912 Homo sapiens chromos
gb_htg2:AC008195	+	39.00	81.94	1.5e+04	116154	! AC008195 Drosophila melanoga
gb_pr3:HSJ1112D6	-	39.00	80.56	1.8e+04	135305	! AL080317 Human DNA sequence
gb_htg4:AC011957	-	39.00	79.40	2.1e+04	153790	! AC011957 Homo sapiens chromc
gb_htg5:AC011702	-	39.00	78.77	2.3e+04	164757	! AC011702 Drosophila melanoga
gb_in2:AC005286	+	39.00	78.28	2.4e+04	173970	! AC005286 Drosophila melanoga
gb_pr2:CNS01DSC	+	39.00	76.64	3.0e+04	208307	! AL121767 Human chromosome 14
gb_pr2:HSZ95098	+	38.00	147.71	3.33	54	! Z95098 H.sapiens mRNA for T cell

```
        2143    4834: contig of 2692 bp in length
        4835         gap of unknown length
        9211:   contig of 4377 bp in length
        9212         gap of unknown length
       25157:   contig of 15946 bp in length
       25158         gap of unknown length
       66208:   contig of 41051 bp in length
       66209         gap of unknown length
      142796:   contig of 76588 bp in length.
FEATURES        Location/Qualifiers
     source     1..142796
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-7024"
                /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT  47989 a  24580 c  24540 g  45686 t     1 others
ORIGIN

alignment_scores:
     Quality:    47.00        Length:    12
     Ratio:       4.273       Gaps:       0
Percent Similarity: 91.667  Percent Identity: 75.000

alignment_block:
US-08-653-294-18 x AC013772/rev     .

Align seg 1/1 to reverse of: AC013772  from: 1  to: 142796

    1 TyrArgLeuAlaIleArgArgIleLeuArgTyr 12
      |||||||||||::|||||| ||||||::::|||
13521 TATAGACTAGCAGTCAGAAGACATTTACTTAAATAT 13486

seq_name: gb_htg2:HSDJ65SC5

seq_documentation_block:
LOCUS       HSDJ65SC5   306999 bp   DNA           HTG       26-NOV-1999
DEFINITION  Homo sapiens chromosome 6 clone RP4-655C5, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL121956.
VERSION     AL121956.2  GI:6469398
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 306999)
AUTHORS     Sims,S.
TITLE       Direct Submission
JOURNAL     Submitted (26-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 27, 1999 this sequence version replaced gi:6066017.
            IMPORTANT:This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues. The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00008   Length: 1036bp
            Contig_ID: 00019   Length: 9963bp
            Contig_ID: 00031   Length: 1518bp
            Contig_ID: 00039   Length: 1498bp
            Contig_ID: 00059   Length: 1840bp
            Contig_ID: 00062   Length: 1083bp
            Contig_ID: 00088   Length: 1408bp
            Contig_ID: 00092   Length: 1477bp
            Contig_ID: 00112   Length: 3742bp
            Contig_ID: 00142   Length: 1888bp
            Contig_ID: 00150   Length: 3371bp
            Contig_ID: 00179   Length: 1299bp
            Contig_ID: 00183   Length: 3722bp
            Contig_ID: 00192   Length: 1260bp
```

```
Contig_ID: 00211   Length: 2606bp
Contig_ID: 00242   Length: 1648bp
Contig_ID: 00248   Length: 4175bp
Contig_ID: 00272   Length: 1781bp
Contig_ID: 00323   Length: 10365bp
Contig_ID: 00355   Length: 1790bp
Contig_ID: 00387   Length: 4242bp
Contig_ID: 00394   Length: 6397bp
Contig_ID: 00404   Length: 1731bp
Contig_ID: 00421   Length: 1266bp
Contig_ID: 00476   Length: 1153bp
Contig_ID: 00481   Length: 980bp
Contig_ID: 00528   Length: 2380bp
Contig_ID: 00555   Length: 2876bp
Contig_ID: 00567   Length: 1458bp
Contig_ID: 00603   Length: 2634bp
Contig_ID: 00605   Length: 1049bp
Contig_ID: 00638   Length: 1196bp
Contig_ID: 00679   Length: 3286bp
Contig_ID: 00695   Length: 4879bp
Contig_ID: 00709   Length: 3049bp
Contig_ID: 00722   Length: 5922bp
Contig_ID: 00725   Length: 1901bp
Contig_ID: 00750   Length: 1420bp
Contig_ID: 00812   Length: 1513bp
Contig_ID: 00892   Length: 1333bp
Contig_ID: 00914   Length: 14732bp
Contig_ID: 00940   Length: 1506bp
Contig_ID: 00964   Length: 1687bp
Contig_ID: 00973   Length: 1044bp
Contig_ID: 00979   Length: 1055bp
Contig_ID: 01017   Length: 3425bp
Contig_ID: 01057   Length: 3223bp
Contig_ID: 01085   Length: 1282bp
Contig_ID: 01125   Length: 3177bp
Contig_ID: 01137   Length: 3871bp
Contig_ID: 01155   Length: 1111bp
Contig_ID: 01212   Length: 1153bp
Contig_ID: 01219   Length: 5468bp
Contig_ID: 01220   Length: 1788bp
Contig_ID: 01238   Length: 3851bp
Contig_ID: 01243   Length: 4555bp
Contig_ID: 01246   Length: 1021bp
Contig_ID: 01256   Length: 1097bp
Contig_ID: 01288   Length: 3286bp
Contig_ID: 01296   Length: 5117bp
Contig_ID: 01338   Length: 2253bp
Contig_ID: 01370   Length: 1884bp
Contig_ID: 01371   Length: 1208bp
Contig_ID: 01374   Length: 6578bp
Contig_ID: 01413   Length: 1372bp
Contig_ID: 01427   Length: 1097bp
Contig_ID: 01430   Length: 2927bp
Contig_ID: 01435   Length: 1190bp
Contig_ID: 01453   Length: 3272bp
Contig_ID: 01487   Length: 1704bp
Contig_ID: 01519   Length: 1071bp
Contig_ID: 01524   Length: 1081bp
Contig_ID: 01530   Length: 2110bp
Contig_ID: 01546   Length: 1126bp
Contig_ID: 01592   Length: 1168bp
Contig_ID: 01630   Length: 8584bp
Contig_ID: 01632   Length: 1827bp
Contig_ID: 01717   Length: 2693bp
Contig_ID: 01740   Length: 1334bp
Contig_ID: 01745   Length: 1861bp
Contig_ID: 01746   Length: 1912bp
Contig_ID: 01776   Length: 2890bp
Contig_ID: 01822   Length: 2552bp
Contig_ID: 01830   Length: 1669bp
Contig_ID: 01853   Length: 1222bp
Contig_ID: 01904   Length: 1994bp
Contig_ID: 01918   Length: 1102bp
```

```
Contig_ID: 01937   Length: 1573bp
Contig_ID: 01944   Length: 1865bp
Contig_ID: 01965   Length: 1096bp
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
FEATURES          Location/Qualifiers
     source        1..306999
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="6"
                   /clone_lib="RPCI-4"
                   /clone="RP4-655C5"
BASE COUNT    72128 a  48318 c  46393 g  68867 t  71293 others
ORIGIN

alignment_scores:
       Quality:        45.00                    Length:       12
       Ratio:          4.091                     Gaps:          0
Percent Similarity:    91.667      Percent Identity: 75.000

alignment_block:
US-08-653-294-18 x HSDJ655C5
Align seg 1/1  to: HSDJ655C5   from: 1  to: 306999

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      ||||||||  ||||||||::||||||::|||
50223 TATAGATTAAATATAAGGAGATTTTACTGAAATAC 50258

seq_name: gb_htg5:AC011145

seq_documentation_block:
LOCUS       AC011145   156243 bp    DNA            HTG       19-NOV-1999
DEFINITION  Homo sapiens clone RP11-2A8, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC011145
VERSION     AC011145.2  GI:6454030
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156243)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2A8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156243)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 19, 1999 this sequence version replaced gi:6006158.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------  Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
```

```
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
            Center project name: L2522
            Center clone name: 2_A_8
            -------------- Summary Statistics
            Sequencing vector: M13; Qualifiers: 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 116065 bases at least Q40
            Consensus quality: 135485 bases at least Q30
            Consensus quality: 148288 bases at least Q20
            Insert size: 164000; agarose-fp
            Insert size: 156243; sum-of-contigs
            Quality coverage: 3.7 in Q20 bases; agarose-fp
            Quality coverage: 3.9 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *          1          4018: contig of 4018 bp in length
 *                           gap of unknown length
 *       4019         18534: contig of 14516 bp in length
 *                           gap of unknown length
 *      18535         29198: contig of 10664 bp in length
 *                           gap of unknown length
 *      29199         44793: contig of 15595 bp in length
 *                           gap of unknown length
 *      44794         65358: contig of 20565 bp in length
 *                           gap of unknown length
 *      65359        103069: contig of 37711 bp in length
 *                           gap of unknown length
 *     103070        156243: contig of 53174 bp in length.
FEATURES          Location/Qualifiers
     source        1..156243
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="RP11-2A8"
                   /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    45948 a  32079 c  31296 g  46894 t  26 others
ORIGIN

alignment_block:
US-08-653-294-18 x AC011145/rev
Align seg 1/1  to reverse of: AC011145   from: 1  to: 156243

    2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      ||||||||||||::::::||||||||||||||||
44664 CGCCTGGCCATAAAAAAGTTTATTGAGATAT 44632

seq_name: gb_htg4:AC012150

seq_documentation_block:
       Quality:        44.00                    Length:       11
       Ratio:          4.000                     Gaps:          0
Percent Similarity:    100.000     Percent Identity: 72.727

LOCUS       AC012150   37459 bp     DNA            HTG       21-OCT-1999
DEFINITION  Homo sapiens chromosome 12p13 clone RPCII-946L16, *** SEQUENCING
            IN PROGRESS ***, 34 unordered pieces.
ACCESSION   AC012150
VERSION     AC012150.1  GI:6091632
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
```

```
ORGANISM  Homo sapiens
          Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37459)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
          Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
          Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
          David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
          Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
          Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
          Guevara,W., Harris,K., Hernandez,J., Hodgson,M., Hogues,M.,
          Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
          Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
          Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
          Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
          Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
          Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
          Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
          Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P.,
          Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
          Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
          Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
TITLE     Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 37459)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (21-OCT-1999) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   * NOTE: This is a 'working draft' sequence. It currently
          * consists of 34 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1:    840: contig of  840 bp in length
          *    841:   1197: contig of  357 bp in length
          *   1198:   2005: contig of  808 bp in length
          *   2006:   2355: contig of  350 bp in length
          *   2356:   2719: contig of  364 bp in length
          *   2720:   3518: contig of  799 bp in length
          *   3519:   4335: contig of  817 bp in length
          *   4336:   4663: contig of  328 bp in length
          *   4664:   5001: contig of  338 bp in length
          *   5002:   6433: contig of 1432 bp in length
          *   6434:   6600: contig of  167 bp in length
          *   6601:   7990: contig of 1390 bp in length
          *   7991:   9023: contig of 1033 bp in length
          *   9024:  10529: contig of 1506 bp in length
          *  10530:  11309: contig of  780 bp in length
          *  11310:  12580: contig of 1271 bp in length
          *  12581:  13673: contig of 1093 bp in length
          *  13674:  15006: contig of 1333 bp in length
          *  15007:  16536: contig of 1530 bp in length
          *  16537:  17561: contig of 1025 bp in length
          *  17562:  19236: contig of 1675 bp in length
          *  19237:  20231: contig of  995 bp in length
          *  20232:  21551: contig of 1320 bp in length
          *  21552:  22756: contig of 1205 bp in length
          *  22757:  23841: contig of 1085 bp in length
          *  23842:  25119: contig of 1278 bp in length
          *  25120:  26287: contig of 1168 bp in length
          *  26288:  28623: contig of 2336 bp in length
          *  28624:  29426: contig of  803 bp in length
          *  29427:  31276: contig of 1850 bp in length
          *  31277:  32391: contig of 1115 bp in length
          *  32392:  33741: contig of 1350 bp in length
          *  33742:  35970: contig of 2229 bp in length
          *  35971:  37459: contig of 1489 bp in length.
FEATURES             Location/Qualifiers
     source          1..37459
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12p13"
                     /clone="RPCI11-946L16"
BASE COUNT    11115 a    7909 c    7207 g   11198 t    30 others
ORIGIN

alignment_scores:
          Quality:    42.00         Length:    12
          Ratio:       3.818        Gaps:       0
Percent Similarity:  91.667     Percent Identity:  58.333

alignment_block:
US-08-653-294-18 x AC012150/rev

Align seg 1/1  to reverse of: AC012150  from: 1  to: 37459

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      |||||||||:::|||::::||| :::|||||||
18153 TATCGCCTAAGTATTAAGCGCAGTATCATTAGATAT 18118

seq_name: gb_in1:DMRNA3

seq_documentation_block:
LOCUS       DMRNA3         1050 bp    DNA            INV    31-MAR-1992
DEFINITION  Three Drosophila melanogaster genes for transfer RNAs (Glu
            specific).
ACCESSION   V00238 J01146
VERSION     V00238.1  GI:8458
KEYWORDS    transfer RNA; transfer RNA-Glu.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1050)
AUTHORS     Hosbach,H.A., Silberklang,M. and McCarthy,B.J.
TITLE       Evolution of a D. melanogaster glutamate tRNA gene cluster
JOURNAL     Cell 21 (1), 169-178 (1980)
MEDLINE     81001852
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            1..1050
                     /gene="tRNA:E4:62Aa"
                     /note="transfer RNA:glu4:62Aa"
                     /allele=""
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            25..96
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            262..333
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            464..535
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
BASE COUNT    324 a    190 c    213 g    323 t
ORIGIN

alignment_scores:
          Quality:    41.00         Length:    12
          Ratio:       3.727        Gaps:       0
Percent Similarity:  91.667     Percent Identity:  58.333

alignment_block:
```

US-08-653-294-18 x DMRNA3
Align seg 1/1   to: DMRNA3   from: 1 to: 1050

```
    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      |||||||:::::|||||| ||||||
  922 TACAGAATGTCGATGCAACGCATTGCTTTACGTTAT 957
```

seq_name: gb_pr2:MMHLAFHOM

seq_documentation_block:
LOCUS   MMHLAFHOM   1056 bp   mRNA   PRI   25-FEB-1993
DEFINITION M.mulatta HLA-F like mRNA.
ACCESSION   Z21819
VERSION   Z21819.1 GI:38568
KEYWORDS   HLA-F gene.
SOURCE   rhesus monkey.
ORGANISM   Macaca mulatta
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE   1 (bases 1 to 1056)
AUTHORS   Bontrop,R.E.
TITLE   Characterization of the rhesus macaque (Macaca mulatta) equivalent
            of HLA-F
JOURNAL   Immunogenetics (1993) In press
REFERENCE   2 (bases 1 to 1056)
AUTHORS   Bontrop,R.R.
TITLE   Direct Submission
JOURNAL   Submitted (16-FEB-1993) Ronald R.E. Bontrop Ph.D, chronic and
            infectious diseases, ITRI-TNO, Lange Kleiweg 151, Rijswijk, 2280
            HV, The netherlands
FEATURES   Location/Qualifiers
     source   1..1056
                     /organism="Macaca mulatta"
                     /isolate="1km"
                     /db_xref="taxon:9544"
     mRNA   <1..>1056
     CDS   10..1056
                     /codon_start=1
                     /product="HLA-F like protein"
                     /protein_id="CAA79885.1"
                     /db_xref="GI:38569"
                     /db_xref="SWISS-PROT:P33617"
                     /translation="MAPRTLLLVLSGALALTETWAGSHSLRYFSTAVSRPGRGEPOYR
                     YIAVSYVDDTQFLRFDSDAAIPRMEPRAPWVEQEGPQYWERTGYAKANARTDRVALR
                     KLLLRYNQSEAGSHTLQGMNGCDMGPDGRLLRGYHQHAYDGKDYISLNEDLRSWTAAD
                     TVARITQRFYEAEEYAEEFRTYLEGECLELLRRYLENGKETLQRADPPKAHLAHHPVS
                     DREATLRCWALGFYPDEITILTWQRDGEEQTQDTELVETRPAGDGTFQKWAAVVPSGE
                     EQRYTCHVQHEGLPQPTLRWESSSQPTIPIVGIVAGLAVLAVVVTGAVVAAVMWRKK
                     SSDRNRGSYSQAAM"
     mat_peptide   76..1053
                     /product="HLA-F like protein"
BASE COUNT   214 a   318 c   349 g   175 t
ORIGIN

alignment_scores:
         Quality: 41.00   Length: 11
         Ratio: 3.727   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x MMHLAFHOM

Align seg 1/1   to: MMHLAFHOM   from: 1 to: 1056

```
    2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      |||:|||:::::|||||| ||||||||
  301 CGAGTGGCCCTGAGGAAGCTGCTCCTCCGCTAC 333
```

seq_name: gb_htg7:AC017966

seq_documentation_block:

LOCUS   AC017966   8849 bp   DNA   *** SEQUENCING HTG   09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017966
VERSION   AC017966.1 GI:6553224
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 8849)
AUTHORS   Adams,M. and Venter,J.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10212817 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES   Location/Qualifiers
     source   1..8849
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT   2604 a   1735 c   1884 g   2626 t
ORIGIN

alignment_scores:
         Quality: 41.00   Length: 12
         Ratio: 3.727   Gaps: 0
Percent Similarity: 91.667   Percent Identity: 58.333

alignment_block:
US-08-653-294-18 x AC017966

Align seg 1/1   to: AC017966   from: 1 to: 8849

```
    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      ||||||:::::::::::||||||| |||||||||
  844 TACAGAATGTCGATGCAACGCATTGCTTTACGTTAT 879
```

seq_name: gb_htg5:AC013927

seq_documentation_block:
LOCUS   AC013927   12771 bp   DNA   *** SEQUENCING HTG   16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013927
VERSION   AC013927.1 GI:6437408
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 12771)
AUTHORS   Adams,M. and Venter,J.C.
TITLE   Direct Submission
JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10213548 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accesion number will be preserved.
FEATURES   Location/Qualifiers
     source   1..12771
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"

```
BASE COUNT      3823 a     2428 c     2497 g     4023 t
ORIGIN

alignment_scores:
     Quality:      41.00              Length:        11
     Ratio:         4.100              Gaps:           0
Percent Similarity:  90.909    Percent Identity:  81.818

alignment_block:
US-08-653-294-18 x AC013927/rev      ..

Align seq 1/1  to reverse of: AC013927  from: 1   to: 12771

      2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
        ||||||||||   |||||||||||||::|||
   2697 CGGCTGGCAATCATGCGAATTTATTGCAATAT 2665

seq_name: gb_bal:MLCB2533

seq_documentation_block:
LOCUS       MLCB2533     40245 bp     DNA           BCT      27-AUG-1999
DEFINITION  Mycobacterium leprae cosmid B2533.
ACCESSION   AL035310
VERSION     AL035310.1   GI:4200258
KEYWORDS    ansP; ATP-dependent RNA helicase; ATP-phosphoribosyl transferase;
            ATPase; hisG 5-methyltetrahydrofolate-homocysteine methyl
            transferase; hisI; L-asparagine permease; metH; mttB;
            phosphoribosyl-AMP cyclohydrolase; prcA; prcB; proteasome alpha
            subunit; proteasome beta subunit; protein translocation system;
            pseudogene; RLEP; sec-independent.
SOURCE      Mycobacterium leprae
  ORGANISM  Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 40245)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE     Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2  (bases 1 to 40245)
  AUTHORS   Hamlin,N. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 40245)
  AUTHORS   James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Mycobacterium leprae sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
            Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,
            Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            France Requests for cosmids should be sent to Karin Eiglmeier
            (kei@pasteur.fr)
COMMENT     Notes:
            The Sanger Centre is funded to complete the sequence of M. leprae
            by the Heiser Program for Research in Leprosy and Tuberculosis of
            The New York Community Trust.
            Work in Paris is supported by the Heiser Trust, the Association
            Francaise Raoul Follereau and the Groupement de Recherches et des
            Etudes des Genomes (GIP-GREG).
            Details of M. leprae sequencing at the Sanger Centre are available
            on the world wide web.
            (URL, http://www.sanger.ac.uk/Projects/)
            CDS are numbered using the following system eg MLCB33.01c. ML (M.
            leprae), cB33 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. All CDS
            over 100 codons have been analysed.  Gene prediction is based on
            positional base preference in codons especially where there is an
            increase in the observed/expected third position G + C.  CAUTION:
            We may not have predicted the correct initiation codon. Where
            possible we choose an initiation codon (atg, gtg, or ttg) which is
            preceded by an upstream ribosome binding site sequence (optimally
            5-13bp before the initiation codon). If this cannot be identified
            we choose the most upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1..40245
                     /organism="Mycobacterium leprae"
                     /db_xref="taxon:1769"
                     /clone="cosmid B2533"
     gene            complement(1..1099)
                     /gene="ansP"
     misc_feature    complement(1..31687)
                     /note="overlap with EMBL:ML017 cosmid B2126 from 1 to
                     31682. There are 16 conflicts between this sequence and
                     ours. In each case our sequence has been checked and is
                     thought to be correct"
     CDS             complement(<1..1099)
                     /gene="ansP"
                     /note="MLCB2533.01c, ansP, probable L-asparagine permease,
                     partial CDS, len: >366 aa; highly similar to many
                     amino-acid permeases e.g. ANSP_SALTY (EMBL:U04851)
                     S.typhimurium AnsP, L-asparagine permease (L-asparagine
                     transport protein) (497 aa), fasta scores; opt: 1508
                     z-score: 1696.8 E(): 0, 61.2% identity in 366 aa overlap.
                     Equivalent to M.tuberculosis Rv2127, ansP (MTCY261.26,
                     85.7% identity in 356 aa overlap). Also similar to
                     M.tuberculosis permease Rv0346c, aroP2 (MTCY13E10.06c,
                     75.1% identity in 345 aa overlap). Probable integral
                     membrane protein, contains PS00218 Amino acid permeases
                     signature. Pfam match to entry PF00324 aa_permeases, amino
                     acid permease. Annotated as ORF TR:Q49801, designated
                     aroP2 in M.leprae cosmid EMBL:U00017"
                     /codon_start=1
                     /transl_table=11
                     /product="putative L-asparagine permease"
                     /protein_id="CAA22915.1"
                     /db_xref="GI:4200259"
                     /translation="MATLAESPEPKSGASRAGVLGEEAGYHKGLKPROLQMIGIGGAI
                     GTGLFLGAGGRLAKAGPGLFLVIAVCGVFVFLILRALGELVLHRPSSGSFVSYAREFF
                     GEKAAYVCGWLYLFLDWAMTAIVDTTAIAYLHRWTIFTALPQWTLALLALAVVVMNL
                     ISVEWFGELEFWAALIKVCALMAFLVVGTIFLGGRYPVDGHNTCLSLWTSHGGLFPTG
                     VAQLIVVSSGVMFAYAAVELVGTAAGETVEPKIMPRAINSVIARIAIFYVGSVILLA
                     LLLPYSAFKASESPVTFSKVGYGAGDLMNIVLTAALSSLNAGLYATGRVMHSIA
                     INGSGPKFTARMSKNGVPYGGILLAAVICLCG"
     repeat_unit     1..1168
                     /note="1168 bp perfect direct repeat"
     misc_feature    complement(2..991)
                     /gene="ansP"
                     /note="Pfam match to entry PF00324 aa_permeases, Amino
                     acid permease, score 245.80, E-value 6e-70"
     unsure          438
                     /gene="ansP"
                     /note="ambiguous base T /G"
     misc_feature    complement(833..925)
                     /gene="ansP"
                     /note="PS00218 Amino acid permeases signature"
     misc_feature    1223
                     /note="conflict: T is TG EMBL:ML017"
     gene            complement(1249..2766)
                     /gene="ansP2"
     CDS             complement(1249..2766)
                     /gene="ansP2"
                     /note="MLCB2533.02c, ansP2, probable L-asparagine
```

```
                (MTCY261.20c, 88.7% identity in 1183 aa overlap).
                Annotated as METH_MYCLE, designated metH2 in M.leprae

alignment_scores:
    Quality:    41.00       Length:        12
    Ratio:       4.556      Gaps:           0
    Percent Similarity: 75.000   Percent Identity: 75.000

alignment_block:
US-08-653-294-18 x MLCB2533    ..
Align seg 1/1  to: MLCB2533   from: 1  to: 40245

     1 TyrArgLeuAlaIleArgArgLeuLeuArgTyr 12
       ||||||||||||  ||||||||
 17534 TATCGATTGGCAAGTCGCAGAATATCTGCACGGTAT 17569


seq_name: gb_bal:U00017

seq_documentation_block:
LOCUS       U00017    42157 bp    DNA            BCT    01-MAR-1994
DEFINITION  Mycobacterium leprae cosmid B2126.
ACCESSION   U00017
VERSION     U00017.1   GI:466994
KEYWORDS    .
SOURCE      Mycobacterium leprae.
ORGANISM    Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1 (bases 1 to 42157)
AUTHORS     Smith,D.R.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 42157)
AUTHORS     Robison,K.
TITLE       Direct Submission
JOURNAL     Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
            School, 200 Longwood Avenue, Boston MA 02115
REFERENCE   3 (bases 1 to 42157)
AUTHORS     Robison,K.
TITLE       Direct Submission
JOURNAL     Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
            School, 200 Longwood Avenue, Boston MA 02115
COMMENT     On Mar 31, 1994 this sequence version replaced gi:414223.
            This sequence data was produced by the Genome Sequencing Center
            located at Collaborative Research Incorporated (1365 Main St.,
            Waltham MA, 02159).  617-487-7979).  Please contact Doug Smith
            (smith@cric.com).  The annotation should be considered
            preliminary and incomplete.
FEATURES    Location/Qualifiers
     source     1..42157
                /organism="Mycobacterium leprae"
                /db_xref="taxon:1769"
     CDS        complement(133..1086)
                /note="match to yigU and yigV E.coli; B2126_C1_183"
                /codon_start=1
                /transl_table=11
                /product="u2126a"
                /protein_id="AAA17191.1"
                /db_xref="GI:467006"
                /translation="MNGCRAVVRACDLLRKIKQHYRRSRTNPDATMSLIDHLTELRTR
                LLISLAAIVVTTIFGFIWYSHSIFGLESLGEWLRRPYCSLPQSARADISPDGQCRLLA
                TAPPDQFMLRIKVGMAAGIVLASPVWFYQLWAFITPGLIYTKERRFTVAFVPAAVLFA
                GGTVLAYLLVLSKALGFLIVGSGVQVTALSGDRYFGFLNLLVVFGVSFEFPLLIVML
                NIAGLLTYQRLKSWRRGLIFAMFVFAAVTTPGSDPFSMTALGAALTVLLELAIQLVRL
                HDKRRVKHEALIADDEASVIEPPSSIPERSYTATRSHDDVT"
     CDS        complement(1132..1398)
                /note="match to yigT E.coli; B2126_C1_182"
                /codon_start=1
                /transl_table=11
                /product="u2126b"
                /protein_id="AAA17190.1"
                /db_xref="GI:467005"
```

```
                permease, len: 505 aa; highly similar to many amino-acid
                permeases e.g. ANSP_SALTY (EMBL:U04851) S.typhimurium
                AnsP, L-asparagine permease (L-asparagine transport
                protein) (497 aa), fasta scores;opt: 1891 z-score: 2218.9
                E(): 0, 58.9% identity in 477 aa overlap. Equivalent to
                M.tuberculosis Rv2127 (MTCY261.26, 83.7% identity in 485
                aa overlap).Also similar to M.tuberculosis permease
                Rv0346c (MTCY13E10.06c, 69.8% identity in 473 aa overlap).
                Probable integral membrane protein, contains PS00218 Amino
                acid permeases signature. Pfam match to entry PF00324
                aa_permeases, amino acid permease. Annotated as ORF
                TR:Q49802, designated lysP in M.leprae cosmid EMBL:U00017"
                /codon_start=1
                /transl_table=11
                /product="putative L-asparagine permease"
                /protein_id="CAA22916.1"
                /db_xref="GI:4200260"
                /translation="MATLAESPEPKSGASRAGVLGEEAGYHKGLKPRQLQMIGGAI
                GTGLFLGAGGRLAKAGPGLFLVYAVCGVFFLLIRALGELVLHRPSSGSFVSYAREFF
                GEKAYAYVVGWLYFLDWAMTAIVDTAIAYLHRWTIFTALPQWTALLALAVVLMWML
                ISVEWFGELEFWAALIKVCALMAFLVVGTIFLGGRYPVDGHNTGLSLWTSHGGLFPTG
                VAPLIVVSSGVMFAYAAVELVGTAAGETVEPKKIMPRAINSVIARIAIFYVGSVILLA
                LLLPYSAFKASESSPFVTFFSKVGFYGAGDLMNIVLTRALSSLNAGLYATGRVMHSIA
                INGSGPKFTARMSKNGVPYGGILLAAVICLCGVALNAFNPGQAFEIVLSVAALGIIAG
                WGTIVLCQLRLHKMAKAGIMRRPRFRMPLAPYSGYLTLAFLFAVLVVMAFDKPIGTWT
                VASLIVIVPALIAGWYSIRKRVMTIARERMGYTGPFPAIANPPVQPSERSHSQNP"
misc_feature    complement(1354..2691)
                /gene="ansP2"
                /note="Pfam match to entry PF00324 aa_permeases, Amino
                acid permease, score 501.80, E-value 5.1e-147"
repeat_unit     1668..2835
                /note="1168 bp perfect direct repeat"
misc_feature    2104
                /gene="ansP2"
                /note="conflict: C is CT in EMBL:ML017"
misc_feature    complement(2500..2592)
                /gene="ansP2"
                /note="PS00218 Amino acid permeases signature"
gene            complement(2903..3856)
                /gene="MLCB2533.03c"
CDS             complement(2903..3856)
                /gene="MLCB2533.03c"
                /note="MLCB2533.03c, hypothetical protein, len: 317 aa;
                similar to M.tuberculosis hypothetical protein Rv2125
                (MTCY261.21) (EMBL:Z97559) (292 aa), fasta scores; opt:
                1648 z-score: 2382.2 E(): 0, 84.1% identity in 290 aa
                overlap. Also some similarity to M.leprae hypothetical
                protein TR:Q49847 (29.7% identity in 279 aa overlap).
                Annotated as ORF TR:Q49797, hypothetical protein in
                M.leprae cosmid EMBL:U00017"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein MLCB2533.03c"
                /protein_id="CAA22917.1"
                /db_xref="GI:4200261"
                /translation="MPPHRAVTRHASSSALKPYADSVTLRDGGPDRGALPELHNTVVV
                AAFEGWNDASDAASGALEHLNAVWEADPIVEIDDEAYYDYQVNRPVIRQVDGVTRELV
                WPAMRISYCRPPGSDRNVLMHGVEPNMWRTFCTELLTIADRLNVDTVVILGALLAD
                TPHTRPVPVSGAAYSPESARRFGLEETRYEGPTGIAGVFQDACVAARIPAVMFWAAVP
                HYVSHPPNPKATVALLRRVEDVLDVEVPLADLPTQAEDWEQAITEIAAEDDELAEYVH
                SLEQRGDAEVDVNDALGKIDGDALAAEFERYLRRRRPGFGR"
gene            4004..7555
                /gene="metH"
CDS             4004..7555
                /gene="metH"
                /note="MLCB2533.04, metH, probable
                5-methyltetrahydrofolate-homocysteine methyltransferase,
                len: 1183 aa; similar to many members of vitamin-B12
                dependent methionine synthase family e.g. METH_ECOLI
                (EMBL:X16584) E.coli metH (1226 aa), fasta scores; opt:
                1617 z-score: 1000.7 E(): 0, 31.6% identity in 1228 aa
                overlap. Equivalent to M.tuberculosis Rv2124c
```

```
CDS
     /translation="MGSLSPWHWVLVLVVVVLLFGAKKLPDAARSLGKSMRIFKSELR
     EMQTENQAQASALETPMQNPTVVQSQRVVPPWSTEQDHTEARPA"
     complement(1471. .1692)
     /codon_start=1
     /transl_table=11
     /product="B2126_C1_181"
     /protein_id="AAA17189.1"
     /db_xref="GI:467004"
     /translation="MSWIFEYYPMRQAHELPDGSFQAVMTYASDAWMTRLVLGFGSAV
     QVQAPEALAYRVRNAAVAALESYQVTAQA"
CDS  complement(1701. .2444)
     /codon_start=1
     /transl_table=11
     /product="B2126_C2_220"
     /protein_id="AAA17199.1"
     /db_xref="GI:467014"
     /translation="MTQLSTRLVRLLNMVPYFQANPRITRAEAAADLGVSAKQLDQDF
     NQLWVCGLPGYGPGDLIDFEFSGDTIEVTFSAGIDRPLQLTSPEAIGLLVALRALANI
     PGVVDPEAVRSAIAKIEAAAVVMGNEATGSVASVDTRPFSESHAVAAVRAAVRNKQAL
     VIDYSASHDTLTSRIVDPIRYLLVGDHSYLEAWSREAEGVRLFRFDRIVVARELDEP
     AAAPETVRRRCQTHRSSMTTRCCRRRRCG"
CDS  complement(2441. .3436)
     /codon_start=1
     /transl_table=11
     /product="B2126_C3_266"
     /protein_id="AAA17208.1"
     /db_xref="GI:467023"
     /translation="MATSKVERLVNLVIALLSTRGYMTAEKIRSSVAGYSDSPTVEAF
     SRMFERDKNELRDLGIPLEVGKVSALDPSEGYRINRDAYALPPVELTPDEAAAVAVAT
     QLWESQELITATQGALLKLRAAGVDIDPLDTPVVIASSGVSSLRGSEDFLSILLSAI
     GSRQAVQFPYRPSRAEPYTMRNVEPWGVITENSCWLLVGHDCDRNATRTFRLSRIGSE
     VAPIGPAGAVTVPDGVLRRIVSDAVAEVSTGATARWVVDGRATALRHAGRPAGVTR
     LGGRDGQVIELDIGSIDRLARDIAGHGADAVVLEPDALRDDVLIRLRAHAGTGPS"
CDS  complement(3443. .3841)
     /codon_start=1
     /transl_table=11
     /product="B2126_C3_265"
     /protein_id="AAA17207.1"
     /db_xref="GI:467022"
     /translation="MDTEIDWVIKRKLFQRYQDRDNMELTDPKIAQLDLAYHDIKRGR
     GVPDLLQRKGLAARVTTDEDIADAVNHPPQTRARLRGEFISAAQAGRDFTVDWVHL
     KLNDQAORTVLCKDPFRAVDERVKRLLIASM"
CDS  complement(3861. .4871)
     /codon_start=1
     /transl_table=11
     /product="B2126_C2_219"
     /protein_id="AAA17198.1"
     /db_xref="GI:467013"
     /translation="MVGWYPWAWLAHRDAQTTSTLKDVQRRIMGIETEFGVTCTFHGH
     SDVLLPFLVTRQLICGAGKVLQTPKAATFCLSQRAEHIWEGVSSATTRSRPIINTRDE
     RRLSPDEVARYLFRRVVSWGRSSNVFLRNGARLYLDVGSHPEYATAECDNLVQLVTHD
     PHADAEKYRRLHVIVGDSNMCETTTMLKVGTAALMLEMVETGVPFRDFSLDNPIRAIR
     EVSHDITGRRPVRLAGGRQASALDIQREYYTRAFEHLQTREPNVQFEQVVDLWGPSAR
     RH"
CDS  complement(6569. .6862)
     /codon_start=1
     /transl_table=11
     /product="B2126_C3_261"
     /protein_id="AAA17206.1"
     /db_xref="GI:467021"
     /translation="MKIADANVLLYSGNTSSEHTGHPCAGSTVRLSCADRIGFGWVPL
     LVFVRPATKMGLVLRTMSSEDAIGQVADWLTGPSAVLMCLTVRHAAFLVKILV"
CDS  complement(6822. .7115)
     /codon_start=1
     /transl_table=11
     /product="B2126_C2_217"
     /protein_id="AAA17197.1"
     /db_xref="GI:467012"
     /translation="MSFKKVFNDVIRDSVLGRFEPHRFTACITDLALSAVNLDWALPI
     AAEFEDEELVRRLSGFGHERRSFLSHRLAPHRLPARIITIHENRRRAAVFR"
CDS  complement(7253. .8050)
     /codon_start=1
```

```
     /transl_table=11
     /product="B2126_C3_260"
     /protein_id="AAA17205.1"
     /db_xref="GI:467020"
     /translation="MSFPYFISPEQAMREQRELARKGIARGRSVVALAYAGGVLFVAE
     NPSRSLQKISELYDRVGFAAAGKFNEFDNLRRGGIQFADRTGYAYDRRDVTGRQLANV
     YAQTLGTIFTEQAKPVEVELCVAEVAHYGETKPPELYRIITYDGSINDEPHFMVMGGTT
     ESIANALKESYAENASLTDALGIAVVALRAGSADAAGSDQPTLGVASLEVAVLDANRP
     RRAFRRIIGSGLEALLREKDSKGSKGAQNPKGARDSKNSKSYGESTD"
CDS  complement(8047. .8922)
     /note="proteasome, beta subunit; B2126_C1_173"
     /codon_start=1
     /transl_table=11
     /product="prcB"
     /protein_id="AAA17188.1"
     /db_xref="GI:467003"
     /translation="MTRSFPDRLPTNLAFPGISVINQSSFVDLLRRQAPELLPVSLGG
     GQSGGGQQLSHGTTIVVLKYPGGVVIAGDRSTOGNMIAGRDVRKYITDYTATGIA
     GIAAVAVEFARLYAVELEHYEKLEGVPLTFAGKVNRLAIMVRSNLTAAMQGLLALPLL
     AGYDIHAPDQSAGRIVSFDAAGGWNIEEEGYQSVGSGSIFAKSSIKKLYSQVSDADS
     ALRVAIEALYDAADDSATGGPDLVRGIYPTAVTIGAEGAAEVTESRIAELAREIIES
     RSRAYTLGSFGGSEK"
CDS  complement(9142. .10686)
     /codon_start=1
     /transl_table=11
     /product="B2126_C1_172"
     /protein_id="AAA17187.1"
     /db_xref="GI:467002"
     /translation="MQRIIGTEVEYGISSPSDPTANPILTSQAVLAYAAAAGIQRVK
     RTRWDYEVESPLRDARGFDLSRSAGPPVVDADEVGAANMILTNGARLYVDHAHPEYS
     APECTDPLDAVIWDKAGERVMEAAARHVASVPGAAKLQLYKNNVDGKGASYGAHENYL
     MSRQTPFSAIIAGLTPGVRSRQRVTGSGRVGIGPAGDEPGFQLSQRSDYIEVEVGLET
     TLKRGIINTRDEPHADADRYRRLHVIVGDANLAETSYLKLGTTALVLDLIEEGPVHG
     IDLTDLTLARPVHAVHAISRDASLRATVTLVDGRELTGLALQRIYLDRVAKLVTDSRDP
     DPRAADVPKTWVHVLDDQLERDPMDCAELLDWPAKLRLLEGFQRENLNWSAPRLHLVD
     LQYSDVRLDKGLYNRLVARGSMKRLVNEHQVLRAVNNPPTDTRAYFRGECLRRFSADI
     AAASWDSVIFDLGGDSLVRIPFLEPLRGSKAHVGALLDSVDSAAELVEQLTTKPVDPG
     KVGLDR"
repeat_region   11019. .11611
     /rpt_family="RLEP"
CDS  complement(11986. .12465)
     /codon_start=1
     /transl_table=11
     /product="B2126_C1_169"
     /protein_id="AAA17186.1"
     /db_xref="GI:467001"
     /translation="MSSAILYLYARLFLLWMCKEEVVAELDDSFYADYLDLNCKYPDV
     QMFFSVLPNIFVAARPSAKTAVIAPVILKGPFHFTGAPALDGAKSTFKHDDEWTDEW
     AHGRSLDALTLQPRHRLQFNIRISDLECEDSRIAALISTDSRRLACDYSLPALSDAA"
CDS  complement(12950. .13234)
     /codon_start=1
     /transl_table=11
     /product="B2126_C3_259"
     /protein_id="AAA17204.1"
     /db_xref="GI:467019"
     /translation="MAVGDDDRHSAGLLSSIDRHDNGWLQOPSSPAASTVQPLPASR
     IPKTSFKPYHVPMLGSRGKASSLTLAALVQAALDPVKQNCTRSRVLFGT"
CDS  13405. .13665
     /codon_start=1
     /transl_table=11
     /product="B2126_F1_17"
     /protein_id="AAA17209.1"
     /db_xref="GI:467024"
     /translation="MVIVKSERVADLLSSVLVVARNTDLGYORFGIRAGMGDAHDTVT
     PSDSRCPGGPNPPVVLLTLVVFNTYWIVATTLIREHYGWR"
CDS  complement(13640. .13921)
     /codon_start=1
     /transl_table=11
     /product="B2126_C3_258"
     /protein_id="AAA17203.1"
     /db_xref="GI:467018"
     /translation="MVAVTTATIYICILSCEVAITKTLRRSALVSAALGEASSIAATG
     SVRGPAFRSSYGSSTVDFIVGLTRFDGGGEDTGAFAVPVRKIIASHSAP"
```

```
                    CDS            14353 ..14763

alignment_scores:
        Quality:      41.00                Length:           12
        Ratio:       4.556                 Gaps:              0
Percent Similarity: 75.000     Percent Identity: 75.000

alignment_block:
US-08-653-294-18 x U00017/rev

Align seg 1/1  to reverse of: U00017  from: 1  to: 42157

     1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
       |||||||||   ||||||||   |||||||
 14153 TATCGATTGGCAAGTCGCAGAATATCTGCACGGTAT 14118

seq_name: gb_htg6:AC007807

seq_documentation_block:
LOCUS       AC007807   130536 bp    DNA             HTG       23-NOV-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR01E04 (D714) RPCI-98
            01.E.4 map 89E-89E strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 99 unordered pieces.
ACCESSION   AC007807
VERSION     AC007807.4  GI:6466499
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 130536)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 130536)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Nov 23, 1999 this sequence version replaced gi:5670524.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 99 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *     1    524: contig of 524 bp in length
            *   525    604: gap of unknown length
            *   605   1174: contig of 570 bp in length
            *  1175   1254: gap of unknown length
            *  1255   1717: contig of 463 bp in length
            *  1718   1797: gap of unknown length
            *  1798   2374: contig of 577 bp in length
            *  2375   2454: gap of unknown length
            *  2455   3026: contig of 572 bp in length
            *  3027   3106: gap of unknown length
            *  3107   3704: contig of 598 bp in length
            *  3705   3784: gap of unknown length
            *  3785   4363: contig of 579 bp in length
            *  4364   4443: gap of unknown length
            *  4444   4935: contig of 492 bp in length
            *  4936   5015: gap of unknown length
            *  5016   5661: contig of 646 bp in length
            *  5662   5741: gap of unknown length
            *  5742   6374: contig of 633 bp in length
            *  6375   6454: gap of unknown length
            *  6455   7007: contig of 553 bp in length
            *  7008   7087: gap of unknown length
            *  7088   7728: contig of 641 bp in length
            *  7729   7808: gap of unknown length
            *  7809   8382: contig of 574 bp in length
            *  8383   8462: gap of unknown length
            *  8463   9034: contig of 572 bp in length
            *  9035   9114: gap of unknown length
            *  9115   9914: contig of 800 bp in length
            *  9915   9994: gap of unknown length
            *  9995  11174: contig of 1180 bp in length
            * 11175  11254: gap of unknown length
            * 11255  12094: contig of 840 bp in length
            * 12095  12174: gap of unknown length
            * 12175  12954: contig of 780 bp in length
            * 12955  13034: gap of unknown length
            * 13035  13830: contig of 796 bp in length
            * 13831  13910: gap of unknown length
            * 13911  15091: contig of 1181 bp in length
            * 15092  15171: gap of unknown length
            * 15172  15977: contig of 806 bp in length
            * 15978  16057: gap of unknown length
            * 16058  16966: contig of 909 bp in length
            * 16967  17046: gap of unknown length
            * 17047  18238: contig of 1192 bp in length
            * 18239  18318: gap of unknown length
            * 18319  18875: contig of 557 bp in length
            * 18876  18955: gap of unknown length
            * 18956  19682: contig of 727 bp in length
            * 19683  19762: gap of unknown length
            * 19763  20523: contig of 761 bp in length
            * 20524  20603: gap of unknown length
            * 20604  21293: contig of 690 bp in length
            * 21294  21373: gap of unknown length
            * 21374  22036: contig of 663 bp in length
            * 22037  22216: gap of unknown length
            * 22217  22906: contig of 790 bp in length
            * 22907  22986: gap of unknown length
            * 22987  23813: contig of 827 bp in length
            * 23814  23893: gap of unknown length
            * 23894  25072: contig of 1179 bp in length
            * 25073  25152: gap of unknown length
            * 25153  26365: contig of 1213 bp in length
            * 26366  26445: gap of unknown length
            * 26446  27654: contig of 1209 bp in length
            * 27655  27734: gap of unknown length
            * 27735  29379: contig of 1645 bp in length
            * 29380  29459: gap of unknown length
            * 29460  30445: contig of 986 bp in length
            * 30446  30525: gap of unknown length
            * 30526  31238: contig of 713 bp in length
            * 31239  31318: gap of unknown length
            * 31319  32321: contig of 1003 bp in length
            * 32322  32401: gap of unknown length
            * 32402  33223: contig of 822 bp in length
            * 33224  33303: gap of unknown length
            * 33304  34356: contig of 1053 bp in length
```

```
alignment_scores:
                Quality:   41.00            Length:      11
                Ratio:      4.100           Gaps:         0
Percent Similarity:  90.909   Percent Identity:  81.818

alignment_block:
US-08-653-294-18 x AC007807
Align seg 1/1 to: AC007807  from: 1 to: 130536

          2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
            |||||||||||  ||||||||||||::||
      83832 CGGCTGGCAATCATGCGGAATTTATTGCAATAT 83864

seq_name: gb_htg4:AC010564

seq_documentation_block:
LOCUS       AC010564   143914 bp    DNA        HTG    16-OCT-1999
DEFINITION  Drosophila melanogaster chromosome 3L/62A1 clone RPCI98-2701, ***
            SEQUENCING IN PROGRESS ***, 89 unordered pieces.
ACCESSION   AC010564
VERSION     AC010564.4  GI:5917942
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 143914)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
            Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
            Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
            Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
            Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
            Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P.,
            Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
            Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 143914)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 22, 1999 this sequence version replaced gi:5916428.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 89 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    831: contig of  831 bp in length
          *      832   1645: gap of unknown length
          *     1646   2519: contig of  874 bp in length
          *     2520   3350: gap of unknown length
          *     3351   4206: contig of  856 bp in length
          *     4207   5469: gap of unknown length
          *     5470   6313: contig of  844 bp in length
          *     6314   7273: gap of unknown length
```

```
*    34357   34436: gap of unknown length
*    34437   35586: contig of 1150 bp in length
*    35587   35666: gap of unknown length
*    35667   36738: contig of 1072 bp in length
*    36739   36818: gap of unknown length
*    36819   37609: contig of  791 bp in length
*    37610   37689: gap of unknown length
*    37690   38360: contig of  671 bp in length
*    38361   38440: gap of unknown length
*    38441   39809: contig of 1369 bp in length
*    39810   39889: gap of unknown length
*    39890   41023: contig of 1134 bp in length
*    41024   41103: gap of unknown length
*    41104   42977: contig of 1874 bp in length
*    42978   43057: gap of unknown length
*    43058   44786: contig of 1729 bp in length
*    44787   44866: gap of unknown length
*    44867   46003: contig of 1137 bp in length
*    46004   46083: gap of unknown length
*    46084   47291: contig of 1208 bp in length
*    47292   47371: gap of unknown length
*    47372   48499: contig of 1128 bp in length
*    48500   48579: gap of unknown length
*    48580   50440: contig of 1861 bp in length
*    50441   50520: gap of unknown length
*    50521   52847: contig of 2327 bp in length
*    52848   52927: gap of unknown length
*    52928   54658: contig of 1731 bp in length
*    54659   54738: gap of unknown length
*    54739   56642: contig of 1904 bp in length
*    56643   56722: gap of unknown length
*    56723   57911: contig of 1189 bp in length
*    57912   57991: gap of unknown length
*    57992   59244: contig of 1253 bp in length
*    59245   59324: gap of unknown length
*    59325   61378: contig of 2054 bp in length
*    61379   61458: gap of unknown length
*    61459   63414: contig of 1956 bp in length
*    63415   63494: gap of unknown length
*    63495   64673: contig of 1179 bp in length
*    64674   64753: gap of unknown length
*    64754   66123: contig of 1370 bp in length
*    66124   66203: gap of unknown length
*    66204   67421: contig of 1218 bp in length
*    67422   67501: gap of unknown length
*    67502   68787: contig of 1286 bp in length
*    68788   68867: gap of unknown length
*    68868   70731: contig of 1864 bp in length
*    70732   70811: gap of unknown length
*    70812   73028: contig of 2217 bp in length
*    73029   73108: gap of unknown length
*    73109   75062: contig of 1954 bp in length
*    75063   75142: gap of unknown length
*    75143   77625: contig of 2483 bp in length
*    77626   77705: gap of unknown length
*    77706   79879: contig of 2174 bp in length
*    79880   79959: gap of unknown length
*    79960   81595: contig of 1636 bp in length
*    81596   81675: gap of unknown length
*    81676   83635: contig of 1960 bp in length
*    83636   83715: gap of unknown length
*    83716   85451: contig of 1736 bp in length
*    85452   85531: gap of unknown length
*    85532   88005: contig of 2474 bp in length
*    88006   88085: gap of unknown length
*    88086   90039: contig of 1954 bp in length
*    90040   90119: gap of unknown length
*    90120   92344: contig of 2225 bp in length
*    92345   92424: gap of unknown length
*    92425   95514: contig of 3090 bp in length
*    95515   95594: gap of unknown length
*    95595   98885: contig of 3291 bp in length
*    98886   98965: gap of unknown length
```

```
*     7274    8101: contig of    828 bp in length
*     8102    8941: contig of    840 bp in length
*     8942   10144: contig of   1203 bp in length
*    10145   11019: contig of    875 bp in length
*    11020   11825: contig of    806 bp in length
*    11826   13085: contig of   1260 bp in length
*    13086   13937: contig of    852 bp in length
*    13938   14752: contig of    815 bp in length
*    14753   15972: contig of   1220 bp in length
*    15973   16571: contig of    599 bp in length
*    16572   17184: contig of    613 bp in length
*    17185   18018: contig of    834 bp in length
*    18019   18837: contig of    819 bp in length
*    18838   20391: contig of   1554 bp in length
*    20392   21252: contig of    861 bp in length
*    21253   21970: contig of    718 bp in length
*    21971   23026: contig of   1056 bp in length
*    23027   24541: contig of   1515 bp in length
*    24542   25726: contig of   1185 bp in length
*    25727   26570: contig of    844 bp in length
*    26571   27681: contig of   1111 bp in length
*    27682   29230: contig of   1549 bp in length
*    29231   30635: contig of   1405 bp in length
*    30636   31757: contig of   1122 bp in length
*    31758   32899: contig of   1142 bp in length
*    32900   34092: contig of   1193 bp in length
*    34093   34928: contig of    836 bp in length
*    34929   35705: contig of    777 bp in length
*    35706   36891: contig of   1186 bp in length
*    36892   38403: contig of   1512 bp in length
*    38404   39938: contig of   1535 bp in length
*    39939   41157: contig of   1219 bp in length
*    41158   42005: contig of    848 bp in length
*    42006   43465: contig of   1460 bp in length
*    43466   44890: contig of   1425 bp in length
*    44891   45926: contig of   1036 bp in length
*    45927   47498: contig of   1572 bp in length
*    47499   49100: contig of   1602 bp in length
*    49101   50371: contig of   1271 bp in length
*    50372   51976: contig of   1605 bp in length
*    51977   54050: contig of   2074 bp in length
*    54051   56027: contig of   1977 bp in length
*    56028   57411: contig of   1384 bp in length
*    57412   58815: contig of   1404 bp in length
*    58816   60132: contig of   1317 bp in length
*    60133   61360: contig of   1228 bp in length
*    61361   63494: contig of   2134 bp in length
*    63495   64634: contig of   1140 bp in length
*    64635   65855: contig of   1221 bp in length
*    65856   66834: contig of    979 bp in length
*    66835   68045: contig of   1211 bp in length
*    68046   68873: contig of    828 bp in length
*    68874   70019: contig of   1146 bp in length
*    70020   72323: contig of   2304 bp in length
*    72324   73759: contig of   1436 bp in length
*    73760   75834: contig of   2075 bp in length
*    75835   78389: contig of   2555 bp in length
*    78390   80840: contig of   2451 bp in length
*    80841   82111: contig of   1271 bp in length
*    82112   84052: contig of   1941 bp in length
*    84053   86110: contig of   2058 bp in length
*    86111   87488: contig of   1378 bp in length
*    87489   89108: contig of   1620 bp in length
*    89109   91457: contig of   2349 bp in length
*    91458   92907: contig of   1450 bp in length
*    92908   95337: contig of   2430 bp in length
*    95338   97095: contig of   1758 bp in length
*    97096  100091: contig of   2996 bp in length
*   100092  102072: contig of   1981 bp in length
*   102073  104497: contig of   2425 bp in length
*   104498  107508: contig of   3011 bp in length
*   107509  109598: contig of   2090 bp in length
*   109599  112405: contig of   2807 bp in length
*   112406  115204: contig of   2799 bp in length
*   115205  118317: contig of   3113 bp in length
*   118318  121885: contig of   3568 bp in length
*   121886  125221: contig of   3336 bp in length
*   125222  129338: contig of   4117 bp in length
*   129339  132811: contig of   3473 bp in length
*   132812  137892: contig of   5081 bp in length
*   137893  143914: contig of   6022 bp in length.

FEATURES             Location/Qualifiers
     source          1..143914
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3L/62A1"
                     /clone="RPCI98-2701"
BASE COUNT    38720 a   33523 c   32072 g   39042 t     557 others
ORIGIN

alignment_scores:
     Quality:       41.00          Length:       12
     Ratio:          3.727          Gaps:          0
     Percent Similarity:  91.667    Percent Identity:  58.333

alignment_block:
US-08-653-294-18 x AC010564

Align seg 1/1  to: AC010564     from:  1  to: 143914

        1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
          |||||:::::::::::::||||||   ||||||||
    58362 TACAGAATGTCGATGCAACGCATTGCTTTACGTTAT 58397

seq_name: gb_htg2:AC005047

seq_documentation_block:
LOCUS       AC005047    159103 bp    DNA            HTG       12-JUN-1998
DEFINITION  Homo sapiens clone RG014E15, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   AC005047
VERSION     AC005047.1  GI:3212944
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159103)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 159103)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     35220: contig of  35220 bp in length
            *    35221     35237: gap of unknown length
            *    35238    159103: contig of 123866 bp in length.
FEATURES             Location/Qualifiers
     source          1..159103
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG014E15"
BASE COUNT    49337 a   31915 c   31970 g   45864 t      17 others
ORIGIN
```

```
alignment_scores:
        Quality:    41.00         Length:         11
        Ratio:      4.100         Gaps:            0
Percent Similarity:  90.909    Percent Identity:  72.727

alignment_block:
US-08-653-294-18 x AC005047      .

Align seq 1/1  to: AC005047   from: 1  to: 159103

        1 TyrArgLeuAlaIleArgArgIleLeuLeuArg 11
          ||||| |||::::||||||||||||||:::
    62132 TACAGGTCTGCAGTAAGGAGGATCTTACTAAAG 62164

seq_name: gb_pr4:AC007687

seq_documentation_block:
LOCUS       AC007687  164655 bp   DNA       PRI     28-OCT-1999
DEFINITION  Homo sapiens 3q26.2-27 BAC RPCI11-419H14 (Roswell Park Cancer
            Institute Human BAC Library) complete sequence.
ACCESSION   AC007687
VERSION     AC007687.16  GI:6137876
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164655)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,D.K., Gorrell,J.H.,
            Gorrell.L.L., Guevara,W., Harris,K., He,X., Hernandez,J.,
            Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E.,
            Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
            Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G.,
            Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
            Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
            Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
            Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M.,
            Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
            Vinson,P., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
            Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
            Yu,W., Zhou,X., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Naylor,S.L., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164655)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1999) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 164655)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 28, 1999 this sequence version replaced gi:6087894.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.
```

```
ANNOTATION OF FEATURES:
    STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
    Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
    Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25.3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT---------

----------- Summary Statistics --------------
Contig length:                                 13071
Phrap values in estimate:                      137060
Average error rate (BCM-Phrap estimate):       0.00042196
Fraction of Phrap values less than 40 :        0.038319
Number of consensus changing edits:            85
Number of N's in consensus :                   0

--------- Consensus changing edits ---------
Position   Original+Context           Edited+Context
22         gtcgagggac(c)taataacttc     gtcgagggac(n)nnnnnnnnn
23         tcgagggacc(t)aataacttcg     tcgagggacc(n)nnnnnnnnn
24         cgagggacct(a)ataacttcgt     cgagggacct(n)nnnnnnnnn
25         gagggaccta(a)taacttcgta     gagggaccta(n)nnnnnnnnn
26         agggacctaa(t)aacttcgtat     agggacctaa(n)nnnnnnnnn
27         gggacctaat(a)acttcgtata     gggacctaat(n)nnnnnnnnn
28         ggacctaata(a)cttcgtatag     ggacctaata(n)nnnnnnnnn
29         gacctaataa(c)ttcgtatagc     gacctaataa(n)nnnnnnnnn
30         acctaataac(t)tcgtatagca     acctaataac(n)nnnnnnnnn
31         cctaataact(t)cgtatagcat     cctaataact(n)nnnnnnnnn
32         ctaataactt(c)gtatagcata     ctaataactt(n)nnnnnnnnn
33         taataacttc(g)tatagcatac     taataacttc(n)nnnnnnnnn
34         aataacttcg(t)atagcataca     aataacttcg(n)nnnnnnnnn
35         ataacttcgt(a)tagcatacat     ataacttcgt(n)nnnnnnnnn
36         taacttcgta(t)agcatacatt     taacttcgta(n)nnnnnnnnn
37         aacttcgtat(a)gcatacatta     aacttcgtat(n)nnnnnnnnn
38         acttcgtata(g)catacattat     acttcgtata(n)nnnnnnnnn
39         cttcgtatag(c)atacattata     cttcgtatag(n)nnnnnnnnn
40         ttcgtatagc(a)tacattatac     ttcgtatagc(n)nnnnnnnnn
41         tcgtatagca(t)acattatacg     tcgtatagca(n)nnnnnnnnn
42         cgtatagcat(a)cattatacga     cgtatagcat(n)nnnnnnnnn
43         gtatagcata(c)attatacgaa     gtatagcata(n)nnnnnnnnn
44         tatagcatac(a)ttatacgaag     tatagcatac(n)nnnnnnnnn
45         atagcataca(t)tatacgaagt     atagcataca(n)nnnnnnnnn
46         tagcatacat(t)atacgaagtt     tagcatacat(n)nnnnnnnnn
47         agcatacatt(a)tacgaagttn     agcatacatt(n)nnnnnnnnn
48         gcatacatta(t)acgaagttnn     gcatacatta(n)nnnnnnnntn
49         catacattat(a)cgaagttnnn     catacattat(n)nnnnnnnntat
50         atacattata(c)gaagttnnnn     atacattata(n)nnnnnnntnnn
51         tacattatac(g)aagttnnnnn     tacattatac(n)nnnnnntatat
52         acattatacg(a)agtnnnnta      acattatacg(a)agtnnnnta
53         cattatacga(a)gttnnnntaa     cattatacga(a)gttnnnntaa
```

```
alignment_scores:
          Quality:    41.00          Length:        11
          Ratio:      4.100          Gaps:          0
Percent Similarity:   90.909   Percent Identity:  72.727

alignment_block:
US-08-653-294-18 x AC007687

Align seg 1/1  to: AC007687  from: 1  to: 164655

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArg 11
      |||  :::|||||||||||||:::::|||||
25920 TACTCTATGCCATCAGGAGGATAATACTGAGA 25952

seq_name: gb_htg1:HSA392M18

seq_documentation_block:
LOCUS       HSA392M18   216149 bp  DNA            HTG   23-NOV-1999
DEFINITION  Homo sapiens chromosome 20 clone RP11-392M18, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL121897
VERSION     AL121897.2  GI:6433871
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216149)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1999) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires: clonerequest@sanger.ac.uk
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 1999 this sequence version replaced gi:6065924.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: bA392M18  Contig_ID: 00029  acc=AL121897
            Length:1432 bp Unfinished: bA392M18  Contig_ID: 00039
            acc=AL121897 Length: 3503 bp Unfinished: bA392M18  Contig_ID:
            00143 acc=AL121897 Length: 1349 bp Unfinished: bA392M18
            Contig_ID: 00174 acc=AL121897 Length: 4959 bp Unfinished:
            bA392M18 Contig_ID: 00213 acc=AL121897 Length: 1677 bp
            Unfinished: bA392M18 Contig_ID: 00233 acc=AL121897 Length: 1233
            bp Unfinished: bA392M18 Contig_ID: 00236 acc=AL121897 Length:
            1104 bp Unfinished: bA392M18 Contig_ID: 00289 acc=AL121897
            Length: 1588 bp Unfinished: bA392M18 Contig_ID: 00342
            acc=AL121897 Length: 1103 bp Unfinished: bA392M18 Contig_ID:
            00363 acc=AL121897 Length: 1045 bp Unfinished: bA392M18
            Contig_ID: 00387 acc=AL121897 Length: 1287 bp Unfinished:
            bA392M18 Contig_ID: 00413 acc=AL121897 Length: 7411 bp
            Unfinished: bA392M18 Contig_ID: 00478 acc=AL121897 Length: 1391
            bp Unfinished: bA392M18 Contig_ID: 00490 acc=AL121897 Length:
            4055 bp Unfinished: bA392M18 Contig_ID: 00523 acc=AL121897
            Length: 1117 bp Unfinished: bA392M18 Contig_ID: 00529
            acc=AL121897 Length: 4870 bp Unfinished: bA392M18 Contig_ID:
            00560 acc=AL121897 Length: 3144 bp Unfinished: bA392M18
            Contig_ID: 00605 acc=AL121897 Length: 2534 bp Unfinished:
            bA392M18 Contig_ID: 00673 acc=AL121897 Length: 1221 bp
            Unfinished: bA392M18 Contig_ID: 00705 acc=AL121897 Length: 1160
            bp Unfinished: bA392M18 Contig_ID: 00800 acc=AL121897 Length:
            1873 bp Unfinished: bA392M18 Contig_ID: 00817 acc=AL121897
            Length: 1940 bp Unfinished: bA392M18 Contig_ID: 00829
            acc=AL121897 Length: 5357 bp Unfinished: bA392M18 Contig_ID:
            00908 acc=AL121897 Length: 1064 bp Unfinished: bA392M18
            Contig_ID: 01023 acc=AL121897 Length: 1331 bp Unfinished:
            bA392M18 Contig_ID: 01114 acc=AL121897 Length: 1386 bp
            Unfinished: bA392M18 Contig_ID: 01166 acc=AL121897 Length: 4885
            bp Unfinished: bA392M18 Contig_ID: 01201 acc=AL121897 Length:
```

```
   54  attatacgaa(g)ttnnnntaag      nnnnnnnnn(n)nntatataag
   55  ttatacgaag(t)tnnnntaagg      nnnnnnnnn(n)ntatataagg
   56  tatacgaagt(t)nnnntaaggg      nnnnnnnnn(n)tatataaggg
   57  atacgaagtt(n)nnntaaggt       nnnnnnnnn(t)tatataaggt
   58  tacgaagttn(n)nntaagggt       nnnnnnnnt(a)tataaggtt
   59  acgaagttnn(n)ntaagggtt       nnnnnnnnta(t)ataaggttc
   60  cgaagttnnn(n)taagggttc       nnnnnnnta(t)ataaggttc
 5843  aaaacaaaa(n)acaaaaatg        aaaaacaaaa(a)acaaaaatgg
10603  tgtgatgca(n)attgctttca       tgtgatgca(g)attgctttca
11108  tatttatctg(t)ctataatctg      tatttatctg(t)ctataatctg
11818  ctcatggtga(n)acagatagtc      ctcatggtga(n)acagatagtc
18266  aaaaaaaac(n)tnttataag        aaaaaaaac(t)ttttataag
18268  aaaaaacnt(n)ttataaagta       aaaaaacnt(t)ttataaagta
23788  cagcagcag(n)tagaaggcag       cagcagcag(n)tagaaggcag
33217  tagtaggta(n)gtacgtgggt       tagtaggta(g)gtacgtgggt
36192  aaaaaaaaa(a)gatatagcaa       aaaaaaaaa(a)gatatagcaa
43205  ttctgatata(n)gcagatacgt      ttctgatata(n)gcagatacgt
56860  agaaatttg(n)tttctttttt       agaaatttg(t)ttttctttttt
56893  tttttttttg(a)ggcggaatct      tttttttttg(a)ggcggaatct
63411  tttctttttc(n)tttttttttt      tttctttttc(n)tttttttttt
64674  ctgtagccct(n)tgattncaga      ctgtagccct(n)tgattccaga
64680  cccttgatt(c)cagagaggga       cccttgatt(c)cagagaggga
64694  agagggagct(n)tgctgccaa       agagggagct(n)tgctgccaa
64702  ctntgctggc(n)aaagcagtnt      ctntgctggc(n)aaagcagttt
64711  cnaaagcagt(n)ttngcactag      caaagcagt(t)ttngcactag
64714  aagcagntt(n)gcactagaaa       aagcagntt(t)gcactagaaa
65683  agctaccac(n)aaaaataaaa       agctaccac(a)aaaaataaaa
68385  taaagagtgt(n)catagataaa      taaagagtgt(t)catagataaa
68505  tcaaatatta(n)ccaaagcatg      tcaaatatta(n)ccaaagcatg
70710  ttttcaaata(n)acttgtctg       ttttcaaata(a)acttgtctg
88521  aaaaaaaaa(g)gagagagaga       aaaaaaaaa(a)gagagagaga
89106  aaaaaaaaaa(a)ccagcactct      aaaaaaaaaa(a)ccagcactct
117652 ataatacaga(n)aaagcaaggc      ataatacaga(a)aaagcaaggc
118982 tatggcctgt(n)tttcctttt       tatggcctgt(t)tttcctttt
120602 ttgacacag(t)taggtgtgt        ttgacacag(t)taggtgtgt
137947 atcagcccc(n)acctggnnag       atcagccccc(c)acctgccag
137954 cccacctgg(c)nagcngcccc       cccaccctgg(c)cagcgcccc
137955 ccnacctgg(n)nagcngcccg       ccnacctgg(c)nagcngcccg
137959 cctgnnagc(n)gccccgnctg       cctgcnagc(c)gccccgtctg
137966 agcngcccc(n)ctgggaggga       agcgccccg(c)ctgggaggga
137979 gggaggggg(n)gggggttag        gggaggggg(c)ggggttag
137987 gngggggggt(n)agccccccgc      ggtggggggt(c)agccccccgc
137997 tagcccccgc(t)cagcagccg       tagcccccgc(c)cagcagccg
138015 cgccccgtc(n)gagaagtatg       cgccccgtc(c)gagaggagg
138020 tcgtcggaga(a)gtatgtgggg      gtccgaga(g)gggaggtggg
138022 gtccgagagg(g)jaggtggggg      cgcagagg(g)gggaggggt
138024 cggagaagta(t)gtgggggggt      gtcagcccc(c)cacctggcca
138042 gtcagcccc(n)nacttgggca       gtcagcccc(c)cacctggcca
138043 gtcagcccc(c)cacctggcca       gtcagcccc(c)cacctggcca
138046 agccccccna(c)ttgggcagcg      agcccccnac(c)ttgggcagcg
138050 ccnnacttgg(g)cagcagccg       cccacctgg(c)cagccgcccc
138055 ctgggcagc(a)gccccgtccg       ctgggccagc(c)gccccgtccg
138067 cccgtccgg(t)aggt.            cccgtccgg(g)aggt.
```

------- Distribution of Quality < 40 Bases ----------

```
1000|
 900|
 800|
 700|
 600|
 500|
 400|                            *  *
 300|                         *  *  *  *
 200|               *  *  *   *  *  *  *
 100|  *  *  *  *   *  *  *   *  *  *  *
   0|  *  *  *  *   *  *  *   *  *  *  *
      --------------------------------------
       5  10  15  20  25  30  35  40
              Phrap Value Range
```

`#` `bases`

```
15088 bp Unfinished: bA392M18  Contig_ID: 01203  acc=AL121897
Length: 6619 bp Unfinished: bA392M18 Contig_ID: 01205
acc=AL121897 Length: 1469 bp Unfinished: bA392M18 Contig_ID:
01209  acc=AL121897 Length: 1040 bp Unfinished: bA392M18
Contig_ID: 01237 acc=AL121897 Length: 1286 bp Unfinished:
bA392M18 Contig_ID: 01270  acc=AL121897 Length: 5408 bp
Unfinished: bA392M18 Contig_ID: 01282 acc=AL121897 Length: 1748
bp Unfinished: bA392M18 Contig_ID: 01335 acc=AL121897 Length:
9850 bp Unfinished: bA392M18 Contig_ID: 01336  acc=AL121897
Length: 4746 bp Unfinished: bA392M18 Contig_ID: 01358
acc=AL121897 Length: 3139 bp Unfinished: bA392M18 Contig_ID:
01428  acc=AL121897 Length: 1701 bp Unfinished: bA392M18
Contig_ID: 01460  acc=AL121897 Length: 3457 bp Unfinished:
bA392M18 Contig_ID: 01489  acc=AL121897 Length: 8480 bp
Unfinished: bA392M18 Contig_ID: 01515  acc=AL121897 Length: 1293
bp Unfinished: bA392M18 Contig_ID: 01517 acc=AL121897 Length:
11602 bp Unfinished: bA392M18 Contig_ID: 01582 acc=AL121897
Length: 4916 bp Unfinished: bA392M18 Contig_ID: 01662
acc=AL121897 Length: 10092 bp Unfinished: bA392M18 Contig_ID:
01664 acc=AL121897 Length: 13254 bp Unfinished: bA392M18
Contig_ID: 01696 acc=AL121897 Length: 1229 bp Unfinished:
bA392M18 Contig_ID: 01731 acc=AL121897 Length: 4821 bp
Unfinished: bA392M18 Contig_ID: 01785 acc=AL121897 Length: 2292
bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES             Location/Qualifiers
     source          1..216149
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-392M18"
                     /clone_lib="RPCI-11.2"
BASE COUNT     47206 a  42210 c  41083 g  48026 t  37624 others
ORIGIN

alignment_scores:
      Quality:     41.00           Length:        12
      Ratio:        3.727          Gaps:           0
Percent Similarity: 91.667    Percent Identity: 66.667

alignment_block:
US-08-653-294-18 x HSA392M18     ..

Align seg 1/1  to: HSA392M18   from: 1  to: 216149

   1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
     :::||||||||::||  |||||||:::|||||:|||||
132492 CACAGGCTTGCTGTAAGGATTATTTTAGTGAGGTAT 132527
```

OM of: US-08-653-294-18 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000   1:28 PM

About: Results were produced by the GenCore software, version 4.5,
      Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
-DB=N_Geneseq_36: -OFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-08-653-294-18
Query length: 12
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 590.520000

score_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
N_Geneseq_36:Q50946		37.00	103.92	345.54	3279	! Sequence encoding protein L. N
N_Geneseq_36:Q51556		37.00	103.92	345.54	3279	! Sequence encoding protein L. N
N_Geneseq_36:T17455		37.00	86.47	3.2e+03	24025	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17515		37.00	86.47	3.2e+03	24025	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T18325		37.00	86.47	3.2e+03	24026	! BRCA1, human breast and ovari
N_Geneseq_36:T17512		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17513		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17514		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17516		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17517		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17518		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17519		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17521		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17522		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17523		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17524		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17526		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17527		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17528		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17529		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17530		37.00	86.47	3.2e+03	24026	! BRCA1 genomic sequence. New n
N_Geneseq_36:T32612		37.00	86.47	3.2e+03	24029	! BRCA1, human breast and ovari
N_Geneseq_36:T17520		37.00	86.47	3.2e+03	24026	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T17525		37.00	86.47	3.2e+03	24031	! Mutated BRCA1 genomic sequenc
N_Geneseq_36:T19381		36.00	129.41	13.15	119	! Human gene signature HUMGS00405
N_Geneseq_36:X04340		36.00	104.84	307.14	1966	! Human secreted protein gene 30
N_Geneseq_36:O23523	+	36.00	104.28	329.69	2094	! ORF3 encoding Abl 105 phage re
N_Geneseq_36:V82467		36.00	102.10	436.11	2686	! Recombinant heat-resistant tre
N_Geneseq_36:V82472		36.00	101.75	456.22	2796	! Recombinant heat-resistant tre
N_Geneseq_36:X20541		36.00	84.75	4.0e+03	19483	! Polynucleotide sequence from
N_Geneseq_36:Q81920	+	35.00	127.54	16.69	98	! Interferon-gamma binding oligonu
N_Geneseq_36:T67384		35.00	120.50	41.20	219	! H. Pylori outer membrane protei
N_Geneseq_36:V44420		35.00	118.86	50.83	264	! Mycobacterium tuberculosis anti
N_Geneseq_36:V64529		35.00	118.86	50.83	264	! M. tuberculosis immunogenic pol
N_Geneseq_36:X30465		35.00	108.98	180.59	816	! H. Pylori outer membrane protei
N_Geneseq_36:X30429		35.00	108.98	180.59	816	! H. Pylori outer membrane protei
N_Geneseq_36:T68162		35.00	108.72	186.57	840	! H. Pylori outer membrane protei
N_Geneseq_36:Q55183		35.00	105.61	278.27	1199	! Sequence for a synthetic lacP-
N_Geneseq_36:Q62660	-	35.00	105.61	278.27	1199	! gp41 HIV envelope protein codi
N_Geneseq_36:N90797		35.00	105.57	279.58	1204	! Synthetic p41 gene with p120 c
N_Geneseq_36:X13394	+	35.00	104.69	313.18	1332	! Enterococcus faecalis genome d

N_Geneseq_36:Q33064   +   35.00   101.17   491.68   1990   ! Encodes Babesia bovis 60kD
N_Geneseq_36:T18995   +   35.00   101.17   491.68   1990   ! Babesia merozoite surface p
N_Geneseq_36:V34005   -   35.00   97.73   764.01   2946   ! Human Rab3-GAP gene. GTP hy
N_Geneseq_36:Q70609   -   35.00   93.51   1.3e+03   4768   ! Hamster HMG-CoA reductase c

seq_name: N_Geneseq_36:Q50946

seq_documentation_block:
ID   Q50946 standard; cDNA; 3279 BP.
AC   Q50946;
DT   18-MAY-1994 (first entry)
DE   Sequence encoding protein L.
KW   Peptide; immunoglobulin; binding; analysis; purification; ELISA;
KW   enzyme linked immunoabsorbant assay; ss.
OS   Peptococcus magnus.
FH   Key   Location/Qualifiers
FT   cds   103..3185
FT   /tag= a
FT   /product= Protein L.
FT   repeat_unit   490..573
FT   /tag= b
FT   /note= "Repeat units are not adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide positions 673 and 856"
FT   repeat_unit   574..672
FT   /tag= c
FT   /note= "Repeat units are not adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide position 757"
FT   repeat_unit   949..1044
FT   /tag= d
FT   /note= "Repeat units are not adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide positions 1162, 1375
FT   and 1597"
FT   repeat_unit   1045..1158
FT   /tag= e
FT   /note= "Repeat units are not adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide positions 1261, 1483
FT   and 1705"
FT   repeat_unit   1822..1938
FT   /tag= f
FT   /note= "Repeat units are not adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide positions 2347, 2545 and
FT   2731"
FT   repeat_unit   2914..2934
FT   /tag= g
FT   /note= "Repeat units are adjacent, repetitions
FT   of this sequence are not 100% homologous and
FT   begin at nucleotide positions 2935, 2953,
FT   2968, 2986, 3001, 3019 and 3034"
PN   WO9322439-A.
PD   11-NOV-1993.
PF   07-MAY-1993; G00950.
PR   07-MAY-1992; GB-009804.
PR   24-DEC-1992; GB-026928.
PA   (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
PI   Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
DR   WPI; 93-368798/46.
DR   P-PSDB; R42203.
PT   New immunoglobulin binding proteins derived from Protein L -
PT   which bind immunoglobulin kappa light chains but not albumin or
PT   cell walls
PS   Disclosure; Figure 1; 28pp; English.
CC   The synthetic immunoglobulin binding proteins derived from protein
CC   L comprise repeated sequences from protein L which bind
CC   immunoglobulin kappa light chains. They can be used in protein
CC   analysis, purification procedures and other biochemical processes e.
CC   g. ELISA. The synthetic molecules are of particular advantage if
CC   they are free of regions in protein L which exhibit albumin and cell
CC   wall binding (The repeat regions commencing at nucleotide numbers

```
CC    1045, 1261, 1483 and 1705).
SQ    Sequence  3279 BP;   1505 A;   481 C;   625 G;   668 T;

alignment_scores:
         Quality:    37.00              Length:   12
         Ratio:       3.083             Gaps:      0
Percent Similarity: 100.000    Percent Identity: 50.000

alignment_block:
US-08-653-294-18 x Q50946
Align seg 1/1  to reverse of: Q50946/rev

      1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
        :::|||||||:::::|||:::|||:::||||||||:::
   2746 TTTCGCCCTGGTTTTCAGCGTGCTTTTAAGATTT 2711

seq_name: N_Geneseq_36:Q51556

seq_documentation_block:
ID    Q51556 standard; cDNA; 3279 BP.
AC    Q51556;
DT    18-MAY-1994 (first entry)
DE    Sequence encoding protein L.
KW    Protein; immunoglobulin; binding; immobilisation; light chains;
KW    antibodies; diagnosis; pharmaceutical; ss.
OS    Peptococcus magnus.
FH    Key              Location/Qualifiers
FT    cds              103..3185
FT                     /*tag= a
FT                     /product= Protein L.
FT    repeat_unit      490..573
FT                     /*tag= b
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 673 and 856"
FT    repeat_unit      574..672
FT                     /*tag= c
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide position 757"
FT    repeat_unit      949..1044
FT                     /*tag= d
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 1162, 1375
FT                     and 1597"
FT    repeat_unit      1045..1158
FT                     /*tag= e
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 1261, 1483
FT                     and 1705"
FT    repeat_unit      1822..1938
FT                     /*tag= f
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 2347 and 2545"
FT    repeat_unit      1939..2007
FT                     /*tag= g
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 2479, 2665
FT                     and 2851"
FT    repeat_unit      2035..2094
FT                     /*tag= h
FT                     /note= "Repeat units are not adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide position 2209"
FT    repeat_unit      2095..2208
FT                     /*tag= i
FT                     /note= "Repeat units are not adjacent, repetitions
```

```
CC                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 2269"
FT    repeat_unit      2914..2934
FT                     /*tag= j
FT                     /note= "Repeat units are adjacent, repetitions
FT                     of this sequence are not 100% homologous and
FT                     begin at nucleotide positions 2935, 2953,
FT                     2968, 2986, 3001, 3019 and 3034"
PN    WO9322438-A.
PD    11-NOV-1993.
PF    07-MAY-1993; G00949.
PR    07-MAY-1992; GB-009804.
PA    (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
PI    Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
DR    WPI; 93-368797/46.
DR    P-PSDB; R43699.
PT    Immunoglobulin binding polypeptide, protein L - used for prodn.
PT    of pharmaceuticals and for immobilising antibodies e.g. on
PT    columns, in diagnostic tests and in assays
PS    Disclosure; Figure 1; 29pp; English.
CC    Protein L forms a complex with immunoglobulin Kappa light chain.
CC    Purified protein can be used as a reagent for immobilising
CC    antibodies e.g. on columns, in diagnostic tests and in assays. It
CC    may also be used in the production of pharmaceuticals.
SQ    Sequence  3279 BP;   1505 A;   480 C;   626 G;   668 T;

alignment_scores:
         Quality:    37.00              Length:   12
         Ratio:       3.083             Gaps:      0
Percent Similarity: 100.000    Percent Identity: 50.000

alignment_block:
US-08-653-294-18 x Q51556
Align seg 1/1  to reverse of: Q51556/rev

      1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
        :::|||||||:::::|||:::|||:::||||||||:::
   2746 TTTCGCCCTGGTTTTCAGCGTGCTTTTAAGATTT 2711

seq_name: N_Geneseq_36:T17455

seq_documentation_block:
ID    T17455 standard; cDNA; 24025 BP.
AC    T17455;
DT    07-OCT-1996 (first entry)
DE    Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.
KW    Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW    antibody production; germline alteration; probe; lesion neoplasia; human;
KW    gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS    Homo sapiens.
FH    Key              Location/Qualifiers
FT    exon             256..355
FT                     /*tag= a
FT                     /note= "exon 1"
FT    intron           356..1512
FT                     /*tag= b
FT                     /note= "intron 1"
FT    misc_feature     1295
FT                     /*tag= c
FT                     /note= "known polymorphic site"
FT    exon             1513..1611
FT                     /*tag= d
FT                     /note= "exon 2"
FT    intron           1612..2206
FT                     /*tag= e
FT                     /note= "intron 2"
FT    misc_feature     1925..1937
FT                     /*tag= f
FT                     /note= "indefinite interval within intron 2"
FT    misc_feature     2141
FT                     /*tag= g
```

```
FT   exon            2207..2260
FT                   /note= "known polymorphic site"
FT                   /*tag= h
FT   intron          2261..2677
FT                   /note= "exon 3"
FT                   /*tag= i
FT   misc_feature    2569..2581
FT                   /note= "intron 3"
FT                   /*tag= j
FT   exon            2678..2788
FT                   /note= "indefinite interval within intron 3"
FT                   /*tag= k
FT   misc_feature    2725
FT                   /note= "exon 4"
FT                   /*tag= l
FT   intron          2789..3328
FT                   /note= "known polymorphic site"
FT                   /*tag= m
FT   misc_feature    3063..3075
FT                   /note= "intron 4"
FT                   /*tag= n
FT   exon            3329..3406
FT                   /note= "indefinite interval within intron 4"
FT                   /*tag= o
FT   intron          3407..3813
FT                   /note= "exon 5"
FT                   /*tag= p
FT   misc_feature    3598..3610
FT                   /note= "intron 5"
FT                   /*tag= q
FT   misc_feature    3653
FT                   /note= "indefinite interval within intron 5"
FT                   /*tag= r
FT   exon            3814..3902
FT                   /note= "known polymorphic site"
FT                   /*tag= s
FT   intron          3903..4224
FT                   /note= "exon 6"
FT                   /*tag= t
FT   mutation        4223
FT                   /note= "intron 6"
FT                   /*tag= u
FT   misc_feature    4076..4088
FT                   /note= "site of 1 nucleotide deletion"
FT                   /*tag= v
FT   exon            4225..4364
FT                   /note= "indefinite interval within intron 6"
FT                   /*tag= w
FT   intron          4365..6571
FT                   /note= "exon 7"
FT                   /*tag= x
FT   misc_feature    4391..4392
FT                   /note= "intron 7"
FT                   /*tag= y
FT   misc_feature    4602..4614
FT                   /note= "known polymorphic site"
FT                   /*tag= z
FT   misc_feature    6538
FT                   /note= "indefinite interval within intron 7"
FT                   /*tag= aa
FT   exon            6572..6677
FT                   /note= "known polymorphic site"
FT                   /*tag= ab
FT   intron          6678..9163
FT                   /note= "exon 8"
FT                   /*tag= ac
FT   misc_feature    6823
FT                   /note= "intron 8"
FT                   /*tag= ad
FT   misc_feature    9106
FT                   /note= "known polymorphic site"
FT                   /*tag= ae
FT   exon            9164..9209
FT                   /*tag= af
FT                   /note= "exon 9"
FT   misc_feature    9207
FT                   /*tag= ag
FT                   /note= "known polymorphic site"
FT   intron          9210..10530
FT                   /*tag= ah
FT                   /note= "intron 9"
FT   misc_feature    9376
FT                   /*tag= ai
FT                   /note= "known polymorphic site"
FT   exon            10531..10607
FT                   /*tag= aj
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= ak
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= al
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= am
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bc
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
```

```
FT   misc_feature
FT                   /tag= bd
FT                   /note= "intron 11"
FT                   15284
FT   misc_feature
FT                   /tag= be
FT                   /note= "known polymorphic site"
FT   exon            15425. .15511
FT                   /tag= bf
FT                   15512. .15952
FT                   /note= "exon 12"
FT   intron
FT                   /tag= bg
FT                   15647. .15659
FT                   /note= "intron 12"
FT   misc_feature
FT                   /tag= bh
FT                   /note= "indefinite interval within intron 12"
FT                   15953. .16126
FT   exon
FT                   /tag= bi
FT                   /note= "known polymorphic site"
FT                   16077
FT   misc_feature
FT                   /tag= bj
FT                   /note= "exon 13"
FT                   16127. .16565
FT   intron
FT                   /tag= bk
FT                   /note= "known polymorphic site"
FT                   16243
FT   misc_feature
FT                   /tag= bl
FT                   /note= "intron 13"
```

```
alignment_scores:
        Quality:   37.00          Length:       11
        Ratio:      3.700         Gaps:          0
Percent Similarity: 90.909   Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x T17455/rev

Align seg 1/1 to reverse of: T17455 from: 1 to: 24025

            2 ArgLeuAlaIleArgIleLeuLeuArgTyr 12
              |||||||||||||||::::::::::::::|| |||
        11313 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11281

seq_name: N_Geneseq_36:T17515
```

```
seq_documentation_block:
ID   T17515  standard; cDNA; 24025 BP.
AC   T17515;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PM15.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   exon           256. .355
FT                   /tag= a
FT                   /note= "exon 1"
FT   intron         356. .1512
FT                   /tag= b
FT                   /note= "intron 1"
FT   misc_feature   1295
FT                   /tag= c
FT                   /note= "known polymorphic site"
FT   exon           1513. .1611
FT                   /tag= d
FT                   /note= "exon 2"
FT   intron         1612. .2206
FT                   /tag= e
FT                   /note= "intron 2"
FT   misc_feature   1925. .1937
FT                   /tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature   2141
FT                   /tag= g
FT                   /note= "known polymorphic site"
```

```
FT   exon           2207. .2260
FT                   /tag= h
FT                   /note= "exon 3"
FT   intron         2261. .2677
FT                   /tag= i
FT                   /note= "intron 3"
FT   misc_feature   2569. .2581
FT                   /tag= j
FT                   /note= "indefinite interval within intron 3"
FT   exon           2678. .2788
FT                   /tag= k
FT                   /note= "exon 4"
FT   misc_feature   2725
FT                   /tag= l
FT                   /note= "known polymorphic site"
FT   intron         2789. .3328
FT                   /tag= m
FT                   /note= "intron 4"
FT   misc_feature   3063. .3075
FT                   /tag= n
FT                   /note= "indefinite interval within intron 4"
FT   exon           3329. .3406
FT                   /tag= o
FT                   /note= "exon 5"
FT   intron         3407. .3813
FT                   /tag= p
FT                   /note= "intron 5"
FT   misc_feature   3598. .3610
FT                   /tag= q
FT                   /note= "indefinite interval within intron 5"
FT   misc_feature   3653
FT                   /tag= r
FT                   /note= "known polymorphic site"
FT   exon           3814. .3902
FT                   /tag= s
FT                   /note= "exon 6"
FT   intron         3903. .4224
FT                   /tag= t
FT                   /note= "intron 6"
FT   misc_feature   4076. .4088
FT                   /tag= u
FT                   /note= "indefinite interval within intron 6"
FT   exon           4225. .4364
FT                   /tag= v
FT                   /note= "exon 7"
FT   intron         4365. .6571
FT                   /tag= w
FT                   /note= "intron 7"
FT   misc_feature   4391. .4392
FT                   /tag= x
FT                   /note= "known polymorphic site"
FT   misc_feature   6538
FT                   /tag= y
FT                   /note= "indefinite interval within intron 7"
FT   misc_feature   6572. .6677
FT                   /tag= z
FT                   /note= "known polymorphic site"
FT   exon           6678. .9163
FT                   /tag= aa
FT                   /note= "exon 8"
FT   intron         6823
FT                   /tag= ab
FT                   /note= "intron 8"
FT   misc_feature   9106
FT                   /tag= ac
FT                   /note= "known polymorphic site"
FT   mutation       9163. .9208
FT                   /tag= ad
FT                   /note= "site of 1 nucleotide deletion at known
FT                   polymorphic site"
FT   exon           9163. .9208
FT                   /tag= ae
FT                   /note= "exon 9"
```

```
FT   exon          3329..3406
FT                 /*tag= j
FT   intron        3407..3813
FT                 /*tag= k
FT                 /note= "n at 3598-3610 represent an indefinite
FT                 interval within the intron"
FT   exon          3814..3902
FT                 /*tag= l
FT   intron        3903..4224
FT                 /*tag= m
FT                 /note= "n at 4076-4088 represent an indefinite
FT                 interval within the intron"
FT   exon          4225..4364
FT                 /*tag= n
FT   intron        4365..6571
FT                 /*tag= o
FT                 /note= "n at 4602-4614 represent an indefinite
FT                 interval within the intron"
FT   exon          6572..6677
FT                 /*tag= p
FT   intron        6678..9163
FT                 /*tag= q
FT   exon          9164..9207
FT                 /*tag= r
FT   intron        9208..10530
FT                 /*tag= s
FT   exon          10531..10607
FT                 /*tag= t
FT   intron        10608..11597
FT                 /*tag= u
FT                 /note= "n at 11383-11396 represent an indefinite
FT                 interval within the intron"
FT   exon          11598..15023
FT                 /*tag= v
FT   intron        15024..15424
FT                 /*tag= w
FT   exon          15425..15511
FT                 /*tag= x
FT   intron        15512..15952
FT                 /*tag= y
FT                 /note= "n at 15647-15659 represent an indefinite
FT                 interval within the intron"
FT   exon          15953..16126
FT                 /*tag= z
FT   intron        16127..16565
FT                 /*tag= aa
FT                 /note= "n at 16370-16382 represent an indefinite
FT                 interval within the intron"
FT   exon          16566..16692
FT                 /*tag= ab
FT   intron        16693..17535
FT                 /*tag= ac
FT                 /note= "n at 17290-17302 represent an indefinite
FT                 interval within the intron"
FT   exon          17536..17726
FT                 /*tag= ad
FT   intron        17727..18416
FT                 /*tag= ae
FT                 /note= "n at 18299-18312 represent an indefinite
FT                 interval within the intron"
FT   exon          18417..18787
FT                 /*tag= af
FT   intron        18788..19298
FT                 /*tag= ag
FT                 /note= "n at 18952-18964 represent an indefinite
FT                 interval within the intron"
FT   exon          19299..19386
FT                 /*tag= ah
FT   intron        19387..20190
FT                 /*tag= ai
FT                 /note= "n at 19887-19899 represent an indefinite
FT                 interval within the intron"
FT   exon          20191..20267
FT                 /*tag= aj
FT   intron        20268..21094
FT                 /*tag= ak
FT                 /note= "n at 20767-20779 represent an indefinite
FT                 interval within the intron"
FT   exon          21095..21135
FT                 /*tag= al
FT   intron        21136..21583
FT                 /*tag= am
FT                 /note= "n at 21341-21353 represent an indefinite
FT                 interval within the intron"
FT   exon          21584..21667
FT                 /*tag= an
FT   intron        21668..22233
FT                 /*tag= ao
FT                 /note= "n at 21921-21933 represent an indefinite
FT                 interval within the intron"
FT   exon          22234..22288
FT                 /*tag= ap
FT   intron        22289..22832
FT                 /*tag= aq
FT                 /note= "n at 22567-22579 represent an indefinite
FT                 interval within the intron"
FT   exon          22833..22906
FT                 /*tag= ar
FT   intron        22907..23287
FT                 /*tag= as
FT                 /note= "n at 23050-23062 represent an indefinite
FT                 interval within the intron"
FT   exon          23288..23348
FT                 /*tag= at
FT   intron        23349..23698
FT                 /*tag= au
FT                 /note= "n at 23580-23592 represent an indefinite
FT                 interval within the intron"
FT   exon          23699..24026
FT                 /*tag= av
FT   misc_feature  2725
FT                 /*tag= aw
FT                 /note= "polymorphic site"
FT   misc_feature  3653
FT                 /*tag= ax
FT                 /note= "polymorphic site"
FT   misc_feature  4391
FT                 /*tag= ay
FT                 /note= "polymorphic site"
FT   misc_feature  4392
FT                 /*tag= az
FT                 /note= "polymorphic site"
FT   misc_feature  6823
FT                 /*tag= ba
FT                 /note= "polymorphic site"
FT   misc_feature  9106
FT                 /*tag= bb
FT                 /note= "polymorphic site"
FT   misc_feature  9207
FT                 /*tag= bc
FT                 /note= "polymorphic site"
FT   misc_feature  9376
FT                 /*tag= bd
FT                 /note= "polymorphic site"
FT   misc_feature  11908
FT                 /*tag= be
FT                 /note= "polymorphic site"
FT   misc_feature  11994
FT                 /*tag= bf
FT                 /note= "polymorphic site"
FT   misc_feature  12952
FT                 /*tag= bg
FT                 /note= "polymorphic site"
FT   misc_feature  13004
FT                 /*tag= bh
FT                 /note= "polymorphic site"
```

```
FT   misc_feature   13009
FT                  /*tag= bi
FT                  /note= "polymorphic site"
FT   misc_feature   13048
FT                  /*tag= bj
FT                  /note= "polymorphic site"
FT   misc_feature   13238
FT                  /*tag= bk
FT                  /note= "polymorphic site"
FT   misc_feature   13448
FT                  /*tag= bl
FT                  /note= "polymorphic site"
FT   misc_feature   13539
FT                  /*tag= bm
FT                  /note= "polymorphic site"
FT   misc_feature   13951
FT                  /*tag= bn
FT                  /note= "polymorphic site"
FT   misc_feature   14041
FT                  /*tag= bo
FT                  /note= "polymorphic site"
FT   misc_feature   14046
FT                  /note= "polymorphic site"

alignment_scores:
             Quality:   37.00          Length:          11
             Ratio:      3.700         Gaps:             0
Percent Similarity:  90.909   Percent Identity:  63.636

alignment_block:
US-08-653-294-18 x T18325/rev  ..

Align seg 1/1  to reverse of: T18325  from:  1  to: 24026

       2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
         ||||||||||||||||:::|||| |||
   11314 AGATTGGCCATCAGAAAACTGGTTCTGANGTAC 11282

seq_name: N_Geneseq_36:T17512

seq_documentation_block:
ID   T17512 standard: cDNA; 24026 BP.
AC   T17512;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PM04.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   exon           256..355
FT                  /*tag= a
FT                  /note= "exon 1"
FT   intron         356..1512
FT                  /*tag= b
FT                  /note= "intron 1"
FT   misc_feature   1295
FT                  /*tag= c
FT                  /note= "known polymorphic site"
FT   exon           1513..1611
FT                  /*tag= d
FT                  /note= "exon 2"
FT   intron         1612..2206
FT                  /*tag= e
FT                  /note= "intron 2"
FT   misc_feature   1925..1937
FT                  /*tag= f
FT                  /note= "indefinite interval within intron 2"
FT   misc_feature   2141
FT                  /*tag= g
FT                  /note= "known polymorphic site"
FT   exon           2207..2260
FT                  /*tag= h
FT                  /note= "exon 3"
FT   intron         2261..2677
FT                  /*tag= i
FT                  /note= "intron 3"
FT   misc_feature   2569..2581
FT                  /*tag= j
FT                  /note= "indefinite interval within intron 3"
FT   exon           2678..2788
FT                  /*tag= k
FT                  /note= "exon 4"
FT   misc_feature   2725
FT                  /*tag= l
FT                  /note= "known polymorphic site"
FT   intron         2789..3328
FT                  /*tag= m
FT                  /note= "intron 4"
FT   misc_feature   3063..3075
FT                  /*tag= n
FT                  /note= "indefinite interval within intron 4"
FT   exon           3329..3406
FT                  /*tag= o
FT                  /note= "exon 5"
FT   intron         3407..3813
FT                  /*tag= p
FT                  /note= "intron 5"
FT   misc_feature   3598..3610
FT                  /*tag= q
FT                  /note= "indefinite interval within intron 5"
FT   misc_feature   3653
FT                  /*tag= r
FT                  /note= "known polymorphic site"
FT   exon           3814..3902
FT                  /*tag= s
FT                  /note= "exon 6"
FT   intron         3903..4224
FT                  /*tag= t
FT                  /note= "intron 6"
FT   misc_feature   4076..4088
FT                  /*tag= u
FT                  /note= "indefinite interval within intron 6"
FT   exon           4225..4364
FT                  /*tag= v
FT                  /note= "exon 7"
FT   intron         4365..6571
FT                  /*tag= w
FT                  /note= "intron 7"
FT   misc_feature   4391..4392
FT                  /*tag= x
FT                  /note= "indefinite interval within intron 7"
FT   misc_feature   4602..4614
FT                  /*tag= y
FT                  /note= "known polymorphic site"
FT   misc_feature   6538
FT                  /*tag= z
FT                  /note= "known polymorphic site"
FT   exon           6572..6677
FT                  /*tag= aa
FT                  /note= "exon 8"
FT   intron         6678..9163
FT                  /*tag= ab
FT                  /note= "intron 8"
FT   misc_feature   6823
FT                  /*tag= ac
FT                  /note= "known polymorphic site"
FT   misc_feature   9106
FT                  /*tag= ad
FT                  /note= "known polymorphic site"
FT   exon           9164..9209
FT                  /*tag= ae
FT                  /note= "exon 9"
FT   misc_feature   9207
FT                  /*tag= af
FT                  /note= "known polymorphic site"
```

```
FT   intron          9210..10530
FT                   /*tag= ag
FT                   /note= "intron 9"
FT   misc_feature    9376
FT                   /*tag= ah
FT                   /note= "known polymorphic site"
FT   exon            10531..10607
FT                   /*tag= ai
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   mutation        15284
FT                   /*tag= bd
FT                   /note= "C to A mutation at known polymorphic site"
FT   exon            15425..15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron          15512..15952
FT                   /*tag= bf
FT                   /note= "intron 12"
FT   misc_feature    15647..15659
FT                   /*tag= bg
FT                   /note= "indefinite interval within intron 12"
FT   exon            15953..16126
FT                   /*tag= bh
FT                   /note= "exon 13"
FT   misc_feature    16077
FT                   /*tag= bi
FT                   /note= "known polymorphic site"
FT   intron          16127..16565
FT                   /*tag= bj
FT                   /note= "intron 13"
FT   misc_feature    16243
FT                   /*tag= bk
FT                   /note= "known polymorphic site"
FT   misc_feature    16370..16382
FT                   /*tag= bl
FT                   /note= "known polymorphic site"

alignment_scores:
             Quality:    37.00             Length:        11
               Ratio:     3.700              Gaps:         0
  Percent Similarity:    90.909   Percent Identity:    63.636

alignment_block:
US-08-653-294-18 x T17512/rev

Align seg 1/1 to reverse of: T17512  from: 1  to: 24026

       2 ArgLeuAlaIleArgIleLeuLeuArgTyr 12
         |||||||||||||||::::::|||  ||
   11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17513

seq_documentation_block:
ID   T17513 standard; cDNA; 24026 BP.
AC   T17513;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PM05.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   exon             256..355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron           356..1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature     1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon             1513..1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron           1612..2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature     1925..1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature     2141
FT                   /*tag= g
FT                   /note= "known polymorphic site"
FT   exon             2207..2260
FT                   /*tag= h
FT                   /note= "exon 3"
```

```
FT   intron          2261..2677
FT                   /*tag= i
FT                   /note= "intron 3"
FT   misc_feature    2569..2581
FT                   /*tag= j
FT                   /note= "indefinite interval within intron 3"
FT   exon            2678..2788
FT                   /*tag= k
FT                   /note= "exon 4"
FT   misc_feature    2725
FT                   /*tag= l
FT                   /note= "known polymorphic site"
FT   intron          2789..3328
FT                   /*tag= m
FT                   /note= "intron 4"
FT   misc_feature    3063..3075
FT                   /*tag= n
FT                   /note= "indefinite interval within intron 4"
FT   exon            3329..3406
FT                   /*tag= o
FT                   /note= "exon 5"
FT   intron          3407..3813
FT                   /*tag= p
FT                   /note= "intron 5"
FT   misc_feature    3598..3610
FT                   /*tag= q
FT                   /note= "indefinite interval within intron 5"
FT   misc_feature    3653
FT                   /*tag= r
FT                   /note= "known polymorphic site"
FT   exon            3814..3902
FT                   /*tag= s
FT                   /note= "exon 6"
FT   intron          3903..4224
FT                   /*tag= t
FT                   /note= "intron 6"
FT   misc_feature    4076..4088
FT                   /*tag= u
FT                   /note= "indefinite interval within intron 6"
FT   exon            4225..4364
FT                   /*tag= v
FT                   /note= "exon 7"
FT   intron          4365..6571
FT                   /*tag= w
FT                   /note= "intron 7"
FT   misc_feature    4391..4392
FT                   /*tag= x
FT                   /note= "known polymorphic site"
FT   misc_feature    4602..4614
FT                   /*tag= y
FT                   /note= "indefinite interval within intron 7"
FT   misc_feature    6538
FT                   /*tag= z
FT                   /note= "known polymorphic site"
FT   exon            6572..6677
FT                   /*tag= aa
FT                   /note= "exon 8"
FT   intron          6678..9163
FT                   /*tag= ab
FT                   /note= "intron 8"
FT   misc_feature    6823
FT                   /*tag= ac
FT                   /note= "known polymorphic site"
FT   misc_feature    9106
FT                   /*tag= ad
FT                   /note= "known polymorphic site"
FT   exon            9164..9209
FT                   /*tag= ae
FT                   /note= "exon 9"
FT   misc_feature    9207
FT                   /*tag= af
FT                   /note= "known polymorphic site"
FT   intron          9210..10530
FT                   /*tag= ag
FT                   /note= "intron 9"
FT   misc_feature    9376
FT                   /*tag= ah
FT                   /note= "known polymorphic site"
FT   exon            10531..10607
FT                   /*tag= ai
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature    15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon            15425..15511
FT                   /*tag= be
```

```
FT   intron
FT                   /note= "exon 12"
FT           15512. .15952
FT                   /*tag= bf
FT   misc_feature
FT                   /note= "intron 12"
FT           15647. .15659
FT                   /*tag= bg
FT   exon
FT                   /note= "indefinite interval within intron 12"
FT           15953. .16126
FT                   /*tag= bh
FT   misc_feature
FT           16077
FT                   /*tag= bi
FT                   /note= "known polymorphic site"
FT           16127. .16565
FT   intron
FT                   /*tag= bj
FT                   /note= "intron 13"
FT           16243
FT   misc_feature
FT                   /*tag= bk
FT           16370. .16382
FT   misc_feature
FT                   /note= "known polymorphic site"
FT                   /*tag= bl

alignment_scores:
        Quality:       37.00                Length:        11
          Ratio:        3.700                 Gaps:         0
Percent Similarity:    90.909    Percent Identity:    63.636

alignment_block:
US-08-653-294-18 x T17513/rev

Align seg 1/1  to reverse of: T17513  from: 1  to: 24026

        2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
          |||||||||||||||:::::::|||  |||
    11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17514

seq_documentation_block:
ID   T17514 standard; cDNA; 24026 BP.
AC   T17514;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PM11.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key          Location/Qualifiers
FT   exon         256. .355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron       356. .1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature 1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon         1513. .1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron       1612. .2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature 1925. .1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature 2141
FT                   /*tag= g
FT                   /note= "known polymorphic site"
FT   exon         2207. .2260
FT                   /*tag= h
FT                   /note= "exon 3"
FT   intron       2261. .2677
```

```
FT   misc_feature
FT                   /*tag= i
FT                   /note= "intron 3"
FT           2569. .2581
FT                   /*tag= j
FT                   /note= "indefinite interval within intron 3"
FT   exon    2678. .2788
FT                   /*tag= k
FT                   /note= "exon 4"
FT   misc_feature 2725
FT                   /*tag= l
FT                   /note= "known polymorphic site"
FT   intron  2789. .3328
FT                   /*tag= m
FT                   /note= "intron 4"
FT   misc_feature 3063. .3075
FT                   /*tag= n
FT                   /note= "indefinite interval within intron 4"
FT   exon    3329. .3406
FT                   /*tag= o
FT                   /note= "exon 5"
FT   intron  3407. .3813
FT                   /*tag= p
FT                   /note= "intron 5"
FT   misc_feature 3598. .3610
FT                   /*tag= q
FT                   /note= "indefinite interval within intron 5"
FT   misc_feature 3653
FT                   /*tag= r
FT                   /note= "known polymorphic site"
FT   exon    3814. .3902
FT                   /*tag= s
FT                   /note= "exon 6"
FT   intron  3903. .4224
FT                   /*tag= t
FT                   /note= "intron 6"
FT   misc_feature 4076. .4088
FT                   /*tag= u
FT                   /note= "indefinite interval within intron 6"
FT   exon    4225. .4364
FT                   /*tag= v
FT                   /note= "exon 7"
FT   intron  4365. .6571
FT                   /*tag= w
FT                   /note= "intron 7"
FT   misc_feature 4391. .4392
FT                   /*tag= x
FT                   /note= "indefinite interval within intron 7"
FT   misc_feature 4602. .4614
FT                   /*tag= y
FT                   /note= "known polymorphic site"
FT   misc_feature 6538
FT                   /*tag= z
FT                   /note= "known polymorphic site"
FT   exon    6572. .6677
FT                   /*tag= aa
FT                   /note= "exon 8"
FT   intron  6678. .9163
FT                   /*tag= ab
FT                   /note= "intron 8"
FT   misc_feature 6823
FT                   /*tag= ac
FT                   /note= "known polymorphic site"
FT   misc_feature 9106
FT                   /*tag= ad
FT                   /note= "known polymorphic site"
FT   exon    9164. .9209
FT                   /*tag= ae
FT                   /note= "exon 9"
FT   misc_feature 9207
FT                   /*tag= af
FT                   /note= "known polymorphic site"
FT   intron  9210. .10530
FT                   /*tag= ag
```

```
FT   misc_feature       9376
FT                      /note= "intron 9"
FT                      /*tag= ah
FT   exon               10531.  .10607
FT                      /note= "known polymorphic site"
FT                      /*tag= ai
FT   intron             10608.  .11597
FT                      /note= "exon 10"
FT                      /*tag= aj
FT   misc_feature       11384.  .11396
FT                      /note= "intron 10"
FT                      /*tag= ak
FT   exon               11598.  .15023
FT                      /note= "indefinite interval within intron 10"
FT                      /*tag= al
FT   misc_feature       11908
FT                      /note= "exon 11"
FT                      /*tag= am
FT   misc_feature       11994
FT                      /note= "known polymorphic site"
FT                      /*tag= an
FT   misc_feature       12952
FT                      /note= "known polymorphic site"
FT                      /*tag= ao
FT   misc_feature       13004
FT                      /note= "known polymorphic site"
FT                      /*tag= ap
FT   misc_feature       13009
FT                      /note= "known polymorphic site"
FT                      /*tag= aq
FT   misc_feature       13048
FT                      /note= "known polymorphic site"
FT                      /*tag= ar
FT   misc_feature       13238
FT                      /note= "known polymorphic site"
FT                      /*tag= as
FT   misc_feature       13448
FT                      /note= "known polymorphic site"
FT                      /*tag= at
FT   misc_feature       13539
FT                      /note= "known polymorphic site"
FT                      /*tag= au
FT   misc_feature       13951
FT                      /note= "known polymorphic site"
FT                      /*tag= av
FT   misc_feature       14041
FT                      /note= "known polymorphic site"
FT                      /*tag= aw
FT   misc_feature       14046
FT                      /note= "known polymorphic site"
FT                      /*tag= ax
FT   misc_feature       14475
FT                      /note= "known polymorphic site"
FT                      /*tag= ay
FT   misc_feature       14874
FT                      /note= "known polymorphic site"
FT                      /*tag= az
FT   misc_feature       14891
FT                      /note= "known polymorphic site"
FT                      /*tag= ba
FT   misc_feature       14966
FT                      /note= "known polymorphic site"
FT                      /*tag= bb
FT   intron             15024.  .15424
FT                      /note= "known polymorphic site"
FT                      /*tag= bc
FT   misc_feature       15284
FT                      /note= "intron 11"
FT                      /*tag= bd
FT   exon               15425.  .15511
FT                      /note= "known polymorphic site"
FT                      /*tag= be
FT                      /note= "exon 12"
```

```
FT   intron             15512.  .15952
FT                      /*tag= bf
FT   misc_feature       15647.  .15659
FT                      /note= "intron 12"
FT                      /*tag= bg
FT   exon               15953.  .16126
FT                      /note= "indefinite interval within intron 12"
FT                      /*tag= bh
FT   misc_feature       16077
FT                      /note= "exon 13"
FT                      /*tag= bi
FT   intron             16127.  .16565
FT                      /note= "known polymorphic site"
FT                      /*tag= bj
FT   misc_feature       16243
FT                      /note= "intron 13"
FT                      /*tag= bk
FT   misc_feature       16370.  .16382
FT                      /note= "known polymorphic site"
FT                      /*tag= bl


alignment_scores:
                  Quality:     37.00        Length:       11
                    Ratio:      3.700         Gaps:        0
   Percent Similarity:  90.909   Percent Identity:   63.636

alignment_block:
   US-08-653-294-18 x T17514/rev  . .

Align seq 1/1  to reverse of: T17514  from: 1  to: 24026

         2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
           ||||||||||||||::|:::::||||  ||
     11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17516

seq_documentation_block:
ID   T17516 standard; cDNA; 24026 BP.
AC   T17516;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PM16.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            256..355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron          356..1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature    1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon            1513..1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron          1612..2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature    1925..1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature    2141
FT                   /*tag= g
FT                   /note= "known polymorphic site"
FT   exon            2207..2260
FT                   /*tag= h
FT                   /note= "exon 3"
FT   intron          2261..2677
FT                   /*tag= 1
```

```
FT   misc_feature
FT                      /note= "intron 3"
FT                      2569..2581
FT                      /*tag= j
FT   exon               "indefinite interval within intron 3"
FT                      2678..2788
FT                      /*tag= k
FT   misc_feature       /note= "exon 4"
FT                      2725
FT                      /*tag= l
FT   intron             /note= "known polymorphic site"
FT                      2789..3328
FT                      /*tag= m
FT   misc_feature       /note= "intron 4"
FT                      3063..3075
FT                      /*tag= n
FT   exon               /note= "indefinite interval within intron 4"
FT                      3329..3406
FT                      /*tag= o
FT   intron             /note= "exon 5"
FT                      3407..3813
FT                      /*tag= p
FT   misc_feature       /note= "intron 5"
FT                      3598..3610
FT                      /*tag= q
FT   misc_feature       /note= "indefinite interval within intron 5"
FT                      3653
FT                      /*tag= r
FT   exon               /note= "known polymorphic site"
FT                      3814..3902
FT                      /*tag= s
FT   intron             /note= "exon 6"
FT                      3903..4224
FT                      /*tag= t
FT   misc_feature       /note= "intron 6"
FT                      4076..4088
FT                      /*tag= u
FT   exon               /note= "indefinite interval within intron 6"
FT                      4225..4364
FT                      /*tag= v
FT   intron             /note= "exon 7"
FT                      4365..6571
FT                      /*tag= w
FT   misc_feature       /note= "intron 7"
FT                      4391..4392
FT                      /*tag= x
FT   misc_feature       /note= "known polymorphic site"
FT                      4602..4614
FT                      /*tag= y
FT   misc_feature       /note= "indefinite interval within intron 7"
FT                      6538
FT                      /*tag= z
FT   exon               /note= "known polymorphic site"
FT                      6572..6677
FT                      /*tag= aa
FT   misc_feature       /note= "exon 8"
FT                      6678..9163
FT                      /*tag= ab
FT   intron             /note= "intron 8"
FT                      6823
FT                      /*tag= ac
FT   misc_feature       /note= "known polymorphic site"
FT                      9106
FT                      /*tag= ad
FT   misc_feature       /note= "known polymorphic site"
FT                      9164..9209
FT                      /*tag= ae
FT   exon               /note= "exon 9"
FT                      9207
FT                      /*tag= af
FT   misc_feature       /note= "known polymorphic site"
FT                      9210..10530
FT                      /*tag= ag
FT   intron             /note= "intron 9"

FT   misc_feature
FT                      9376
FT                      /*tag= ah
FT   exon               /note= "known polymorphic site"
FT                      10531..10607
FT                      /*tag= ai
FT   intron             /note= "exon 10"
FT                      10608..11597
FT                      /*tag= aj
FT   misc_feature       /note= "intron 10"
FT                      11384..11396
FT                      /*tag= ak
FT   exon               /note= "indefinite interval within intron 10"
FT                      11598..15023
FT                      /*tag= al
FT   misc_feature       /note= "exon 11"
FT                      11908
FT                      /*tag= am
FT   misc_feature       /note= "known polymorphic site"
FT                      11994
FT                      /*tag= an
FT   misc_feature       /note= "known polymorphic site"
FT                      12952
FT                      /*tag= ao
FT   misc_feature       /note= "known polymorphic site"
FT                      13004
FT                      /*tag= ap
FT   misc_feature       /note= "known polymorphic site"
FT                      13009
FT                      /*tag= aq
FT   misc_feature       /note= "known polymorphic site"
FT                      13048
FT                      /*tag= ar
FT   misc_feature       /note= "known polymorphic site"
FT                      13238
FT                      /*tag= as
FT   misc_feature       /note= "known polymorphic site"
FT                      13448
FT                      /*tag= at
FT   misc_feature       /note= "known polymorphic site"
FT                      13539
FT                      /*tag= au
FT   misc_feature       /note= "known polymorphic site"
FT                      13951
FT                      /*tag= av
FT   misc_feature       /note= "known polymorphic site"
FT                      14041
FT                      /*tag= aw
FT   misc_feature       /note= "known polymorphic site"
FT                      14046
FT                      /*tag= ax
FT   misc_feature       /note= "known polymorphic site"
FT                      14475
FT                      /*tag= ay
FT   misc_feature       /note= "known polymorphic site"
FT                      14874
FT                      /*tag= az
FT   misc_feature       /note= "known polymorphic site"
FT                      14891
FT                      /*tag= ba
FT   misc_feature       /note= "known polymorphic site"
FT                      14966
FT                      /*tag= bb
FT   intron             /note= "known polymorphic site"
FT                      15024..15424
FT                      /*tag= bc
FT   misc_feature       /note= "intron 11"
FT                      15284
FT                      /*tag= bd
FT   exon               /note= "known polymorphic site"
FT                      15425..15511
FT                      /*tag= be
FT   intron             /note= "exon 12"
FT                      15512..15952
```

```
FT                  /*tag= bf
FT                  /note= "intron 12"
FT   misc_feature   15647..15659
FT                  /*tag= bg
FT                  /note= "indefinite interval within intron 12"
FT   exon           15953..16126
FT                  /*tag= bh
FT                  /note= "exon 13"
FT   misc_feature   16077
FT                  /*tag= bi
FT                  /note= "known polymorphic site"
FT   intron         16127..16565
FT                  /*tag= bj
FT                  /note= "intron 13"
FT   misc_feature   16243
FT                  /*tag= bk
FT                  /note= "known polymorphic site"
FT   misc_feature   16370..16382
FT                  /*tag= bl
FT                  /note= "intron 13"

alignment_scores:
                Quality:  37.00              Length:      11
                  Ratio:  3.700              Gaps:         0
Percent Similarity: 90.909   Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x T17516/rev

Align seq 1/1  to reverse of: T17516  from: 1  to: 24026

        2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
          ||||||||||||||:::::::|| |||
    11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17517

seq_documentation_block:
ID   T17517 standard; cDNA; 24026 BP.
AC   T17517;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PMA02.1.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration: probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key      Location/Qualifiers
FT   exon     256..355
FT                  /*tag= a
FT                  /note= "exon 1"
FT   intron   355..1512
FT                  /*tag= b
FT                  /note= "intron 1"
FT   mutation 1295
FT                  /*tag= c
FT                  /note= "G to A mutation at known polymorphic site"
FT   exon     1513..1611
FT                  /*tag= d
FT                  /note= "exon 2"
FT   intron   1612..2206
FT                  /*tag= e
FT                  /note= "intron 2"
FT   misc_feature 1925..1937
FT                  /*tag= f
FT                  /note= "indefinite interval within intron 2"
FT   misc_feature 2141
FT                  /*tag= g
FT                  /note= "known polymorphic site"
FT   exon     2207..2260
FT                  /*tag= h
FT                  /note= "exon 3"
FT   intron   2261..2677
FT                  /*tag= i
FT                  /note= "intron 3"


FT   misc_feature   2569..2581
FT                  /*tag= j
FT                  /note= "indefinite interval within intron 3"
FT   exon           2678..2788
FT                  /*tag= k
FT                  /note= "exon 4"
FT   misc_feature   2725
FT                  /*tag= l
FT                  /note= "known polymorphic site"
FT   intron         2789..3328
FT                  /*tag= m
FT                  /note= "intron 4"
FT   misc_feature   3063..3075
FT                  /*tag= n
FT                  /note= "indefinite interval within intron 4"
FT   exon           3329..3406
FT                  /*tag= o
FT                  /note= "exon 5"
FT   intron         3407..3813
FT                  /*tag= p
FT                  /note= "intron 5"
FT   misc_feature   3598..3610
FT                  /*tag= q
FT                  /note= "indefinite interval within intron 5"
FT   misc_feature   3653
FT                  /*tag= r
FT                  /note= "known polymorphic site"
FT   exon           3814..3902
FT                  /*tag= s
FT                  /note= "exon 6"
FT   intron         3903..4224
FT                  /*tag= t
FT                  /note= "intron 6"
FT   misc_feature   4076..4088
FT                  /*tag= u
FT                  /note= "indefinite interval within intron 6"
FT   exon           4225..4364
FT                  /*tag= v
FT                  /note= "exon 7"
FT   intron         4365..6571
FT                  /*tag= w
FT                  /note= "intron 7"
FT   misc_feature   4391..4392
FT                  /*tag= x
FT                  /note= "known polymorphic site"
FT   misc_feature   4602..4614
FT                  /*tag= y
FT                  /note= "indefinite interval within intron 7"
FT   misc_feature   6538
FT                  /*tag= z
FT                  /note= "known polymorphic site"
FT   exon           6572..6677
FT                  /*tag= aa
FT                  /note= "exon 8"
FT   intron         6678..9163
FT                  /*tag= ab
FT                  /note= "intron 8"
FT   misc_feature   6823
FT                  /*tag= ac
FT                  /note= "known polymorphic site"
FT   misc_feature   9106
FT                  /*tag= ad
FT                  /note= "known polymorphic site"
FT   exon           9164..9209
FT                  /*tag= ae
FT                  /note= "exon 9"
FT   misc_feature   9207
FT                  /*tag= af
FT                  /note= "known polymorphic site"
FT   intron         9210..10530
FT                  /*tag= ag
FT                  /note= "intron 9"
FT   misc_feature   9376
```

```
FT                   /*tag= ah
FT                   /note= "known polymorphic site"
FT   exon           10531. .10607
FT                   /*tag= ai
FT                   /note= "exon 10"
FT   intron         10608. .11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature   11384. .11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon           11598. .15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature   11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature   11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature   12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature   13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature   13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature   13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature   13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature   13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature   13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature   13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature   14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature   14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature   14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature   14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature   14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature   14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron         15024. .15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature   15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon           15425. .15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron         15512. .15952
FT                   /*tag= bf
```

```
FT                   /note= "intron 12"
FT   misc_feature   15647. .15659
FT                   /*tag= bg
FT                   /note= "indefinite interval within intron 12"
FT   exon           15953. .16126
FT                   /*tag= bh
FT                   /note= "exon 13"
FT   misc_feature   16077
FT                   /*tag= bi
FT                   /note= "known polymorphic site"
FT   intron         16127. .16565
FT                   /*tag= bj
FT                   /note= "intron 13"
FT   misc_feature   16243
FT                   /*tag= bk
FT                   /note= "known polymorphic site"
FT   misc_feature   16370. .16382
FT                   /*tag= bl

alignment_scores:
             Quality:    37.00          Length:        11
             Ratio:      3.700          Gaps:           0
 Percent Similarity:    90.909  Percent Identity:  63.636

alignment_block:
US-08-653-294-18 x T17517/rev

Align seg 1/1  to reverse of: T17517  from:  1  to: 24026

          2 ArgLeuAlaIleArgIleLeuLeuArgTyr 12
            |||||||||||||||::::::::||| |||
      11314 AGATTGGCCATCAGAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17518

seq_documentation_block:
ID   T17518 standard; cDNA; 24026 BP.
AC   T17518;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PMA03.1.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon           256. .355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron         356. .1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature   1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon           1513. .1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron         1612. .2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature   1925. .1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   mutation       2141
FT                   /*tag= g
FT                   /note= "G to C mutation at known polymorphic site"
FT   exon           2207. .2260
FT                   /*tag= h
FT                   /note= "exon 3"
FT   intron         2261. .2677
FT                   /*tag= i
FT                   /note= "intron 3"
FT   misc_feature   2569. .2581
```

```
FT
FT
FT   exon            2678..2788
FT                   /*tag= j
FT                   /note= "indefinite interval within intron 3"
FT                   /*tag= k
FT   misc_feature    2725
FT                   /note= "exon 4"
FT                   /*tag= l
FT                   /note= "known polymorphic site"
FT   intron          2789..3328
FT                   /*tag= m
FT                   /note= "intron 4"
FT   misc_feature    3063..3075
FT                   /*tag= n
FT                   /note= "indefinite interval within intron 4"
FT   exon            3329..3406
FT                   /*tag= o
FT                   /note= "exon 5"
FT   intron          3407..3813
FT                   /*tag= p
FT                   /note= "intron 5"
FT   misc_feature    3598..3610
FT                   /*tag= q
FT                   /note= "indefinite interval within intron 5"
FT   misc_feature    3653
FT                   /*tag= r
FT                   /note= "known polymorphic site"
FT   exon            3814..3902
FT                   /*tag= s
FT                   /note= "exon 6"
FT   intron          3903..4224
FT                   /*tag= t
FT                   /note= "intron 6"
FT   misc_feature    4076..4088
FT                   /*tag= u
FT                   /note= "indefinite interval within intron 6"
FT   exon            4225..4364
FT                   /*tag= v
FT                   /note= "exon 7"
FT   intron          4365..6571
FT                   /*tag= w
FT                   /note= "intron 7"
FT   misc_feature    4391..4392
FT                   /*tag= x
FT                   /note= "known polymorphic site"
FT   misc_feature    4602..4614
FT                   /*tag= y
FT                   /note= "indefinite interval within intron 7"
FT   misc_feature    6538
FT                   /*tag= z
FT                   /note= "known polymorphic site"
FT   exon            6572..6677
FT                   /*tag= aa
FT                   /note= "exon 8"
FT   intron          6678..9163
FT                   /*tag= ab
FT                   /note= "intron 8"
FT   misc_feature    6823
FT                   /*tag= ac
FT                   /note= "known polymorphic site"
FT   misc_feature    9106
FT                   /*tag= ad
FT                   /note= "known polymorphic site"
FT   exon            9164..9209
FT                   /*tag= ae
FT                   /note= "exon 9"
FT   misc_feature    9207
FT                   /*tag= af
FT                   /note= "known polymorphic site"
FT   intron          9210..10530
FT                   /*tag= ag
FT                   /note= "intron 9"
FT   misc_feature    9376
FT                   /*tag= ah

FT   exon            10531..10607
FT                   /*tag= ai
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature    15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon            15425..15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron          15512..15952
FT                   /*tag= bf
FT                   /note= "intron 12"
```

```
FT   misc_feature   15647..15659
FT                  /*tag= bg
FT                  /note= "indefinite interval within intron 12"
FT   exon           15953..16126
FT                  /*tag= bh
FT                  /note= "exon 13"
FT   misc_feature   16077
FT                  /*tag= bi
FT                  /note= "known polymorphic site"
FT   intron         16127..16565
FT                  /*tag= bj
FT                  /note= "intron 13"
FT   misc_feature   16243
FT                  /*tag= bk
FT                  /note= "known polymorphic site"
FT   misc_feature   16370..16382
FT                  /*tag= bl

alignment_scores:
    Quality:             37.00        Length:            11
    Ratio:               3.700        Gaps:              0
Percent Similarity:     90.909   Percent Identity:  63.636

alignment_block:
us-08-653-294-18 x T17518/rev

Align seg 1/1  to reverse of: T17518  from: 1   to: 24026

      2 ArgLeuAlaIleArgArgIleLeuLeuArgArgTyr 12
        |||||||||||||||||:::::::|||  |||
  11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Genseq_36:T17519

seq_documentation_block:
ID   T17519 standard; cDNA; 24026 BP.
AC   T17519;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PMA06.1.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon           256..355
FT                  /*tag= a
FT                  /note= "exon 1"
FT   intron         356..1512
FT                  /*tag= b
FT                  /note= "intron 1"
FT   misc_feature   1295
FT                  /*tag= c
FT                  /note= "known polymorphic site"
FT   exon           1513..1611
FT                  /*tag= d
FT                  /note= "exon 2"
FT   intron         1612..2206
FT                  /*tag= e
FT                  /note= "intron 2"
FT   misc_feature   1925..1937
FT                  /*tag= f
FT                  /note= "indefinite interval within intron 2"
FT   misc_feature   2141
FT                  /*tag= g
FT                  /note= "known polymorphic site"
FT   exon           2207..2260
FT                  /*tag= h
FT                  /note= "exon 3"
FT   intron         2261..2677
FT                  /*tag= i
FT                  /note= "intron 3"
FT   misc_feature   2569..2581
FT                  /*tag= j
FT                  /note= "indefinite interval within intron 3"
FT   exon           2678..2788
FT                  /*tag= k
FT                  /note= "exon 4"
FT   misc_feature   2725
FT                  /*tag= l
FT                  /note= "known polymorphic site"
FT   intron         2789..3328
FT                  /*tag= m
FT                  /note= "intron 4"
FT   misc_feature   3063..3075
FT                  /*tag= n
FT                  /note= "indefinite interval within intron 4"
FT   exon           3329..3406
FT                  /*tag= o
FT                  /note= "exon 5"
FT   intron         3407..3813
FT                  /*tag= p
FT                  /note= "intron 5"
FT   misc_feature   3598..3610
FT                  /*tag= q
FT                  /note= "indefinite interval within intron 5"
FT   mutation       3653
FT                  /*tag= r
FT                  /note= "A to G mutation at known polymorphic site"
FT   exon           3814..3902
FT                  /*tag= s
FT                  /note= "exon 6"
FT   intron         3903..4224
FT                  /*tag= t
FT                  /note= "intron 6"
FT   misc_feature   4076..4088
FT                  /*tag= u
FT                  /note= "indefinite interval within intron 6"
FT   exon           4225..4364
FT                  /*tag= v
FT                  /note= "exon 7"
FT   intron         4365..6571
FT                  /*tag= w
FT                  /note= "intron 7"
FT   misc_feature   4391..4392
FT                  /*tag= x
FT                  /note= "indefinite interval within intron 7"
FT   misc_feature   4602..4614
FT                  /*tag= y
FT                  /note= "known polymorphic site"
FT   misc_feature   6538
FT                  /*tag= z
FT                  /note= "known polymorphic site"
FT   exon           6572..6677
FT                  /*tag= aa
FT                  /note= "exon 8"
FT   intron         6678..9163
FT                  /*tag= ab
FT                  /note= "intron 8"
FT   misc_feature   6823
FT                  /*tag= ac
FT                  /note= "known polymorphic site"
FT   misc_feature   9106
FT                  /*tag= ad
FT                  /note= "known polymorphic site"
FT   exon           9164..9209
FT                  /*tag= ae
FT                  /note= "exon 9"
FT   misc_feature   9207
FT                  /*tag= af
FT                  /note= "known polymorphic site"
FT   intron         9210..10530
FT                  /*tag= ag
FT                  /note= "intron 9"
FT   misc_feature   9376
FT                  /*tag= ah
FT                  /note= "known polymorphic site"
```

```
FT   exon            10531. .10607
FT                   /*tag= ai
FT                   /note= "exon 10"
FT   intron          10608. .11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature    11384. .11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598. .15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024. .15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature    15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon            15425. .15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron          15512. .15952
FT                   /*tag= bf
FT                   /note= "intron 12"
FT   misc_feature    15647. .15659
FT                   /*tag= bg
FT                   /note= "indefinite interval within intron 12"
FT   exon            15953. .16126
FT                   /*tag= bh
FT                   /note= "exon 13"
FT   misc_feature    16077
FT                   /*tag= bi
FT                   /note= "known polymorphic site"
FT   intron          16127. .16565
FT                   /*tag= bj
FT                   /note= "intron 13"
FT   misc_feature    16243
FT                   /*tag= bk
FT                   /note= "known polymorphic site"
FT   misc_feature    16370. .16382
FT                   /*tag= bl
```

```
alignment_scores:
          Quality:   37.00              Length:       11
            Ratio:   3.700                Gaps:        0
 Percent Similarity:  90.909     Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x T17519/rev

Align seg 1/1  to reverse of: T17519  from: 1   to: 24026

      2 ArgLeuAlaIleArgIleLeuLeuArgTyr 12
        |||||||||||||||||::::::::|||  |||
  11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282
```

```
seq_name: N_Geneseq_36:T17521

seq_documentation_block:
ID   T17521 standard; cDNA; 24026 BP.
AC   T17521;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PMA08.1.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            256. .355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron          356. .1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature    1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon            1513. .1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron          1612. .2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature    1925. .1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature    2141
FT                   /*tag= g
FT                   /note= "known polymorphic site"
FT   exon            2207. .2260
FT                   /*tag= h
FT                   /note= "exon 3"
FT   intron          2261. .2677
FT                   /*tag= i
FT                   /note= "intron 3"
FT   misc_feature    2569. .2581
FT                   /*tag= j
FT                   /note= "indefinite interval within intron 3"
```

```
FT   exon            2678..2788
FT                   /*tag= k
FT                   /note= "exon 4"
FT   misc_feature    2725
FT                   /*tag= l
FT                   /note= "known polymorphic site"
FT   intron          2789..3328
FT                   /*tag= m
FT                   /note= "intron 4"
FT   misc_feature    3063..3075
FT                   /*tag= n
FT                   /note= "indefinite interval within intron 4"
FT   exon            3329..3406
FT                   /*tag= o
FT                   /note= "exon 5"
FT   intron          3407..3813
FT                   /*tag= p
FT                   /note= "intron 5"
FT   misc_feature    3598..3610
FT                   /*tag= q
FT                   /note= "indefinite interval within intron 5"
FT   misc_feature    3653
FT                   /*tag= r
FT                   /note= "known polymorphic site"
FT   exon            3814..3902
FT                   /*tag= s
FT                   /note= "exon 6"
FT   intron          3903..4224
FT                   /*tag= t
FT                   /note= "intron 6"
FT   misc_feature    4076..4088
FT                   /*tag= u
FT                   /note= "indefinite interval within intron 6"
FT   exon            4225..4364
FT                   /*tag= v
FT                   /note= "exon 7"
FT   intron          4365..6571
FT                   /*tag= w
FT                   /note= "intron 7"
FT   misc_feature    4391..4392
FT                   /*tag= x
FT                   /note= "indefinite interval within intron 7"
FT   misc_feature    4602..4614
FT                   /*tag= y
FT                   /note= "known polymorphic site"
FT   mutation        6538
FT                   /*tag= z
FT                   /note= "C to T mutation at known polymorphic site"
FT   exon            6572..6677
FT                   /*tag= aa
FT                   /note= "exon 8"
FT   intron          6678..9163
FT                   /*tag= ab
FT                   /note= "intron 8"
FT   misc_feature    9106
FT                   /*tag= ac
FT                   /note= "known polymorphic site"
FT   exon            9164..9209
FT                   /*tag= ad
FT                   /note= "exon 9"
FT   misc_feature    9207
FT                   /*tag= ae
FT                   /note= "known polymorphic site"
FT   intron          9210..10530
FT                   /*tag= af
FT                   /note= "intron 9"
FT   misc_feature    9376
FT                   /*tag= ag
FT                   /note= "known polymorphic site"
FT   exon            10531..10607
FT                   /*tag= ah
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= ai
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= aj
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= ak
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13448
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature    15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon            15425..15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron          15512..15952
FT                   /*tag= bf
FT                   /note= "intron 12"
FT   misc_feature    15647..15659
FT                   /*tag= bg
```

```
FT  exon            15953..16126
FT                  /note= "indefinite interval within intron 12"
FT                  /*tag= bh
FT  misc_feature    16077
FT                  /note= "exon 13"
FT                  /*tag= bi
FT  intron          16127..16565
FT                  /note= "known polymorphic site"
FT                  /*tag= bj
FT  misc_feature    16243
FT                  /note= "intron 13"
FT                  /*tag= bk
FT  misc_feature    16370..16382
FT                  /note= "known polymorphic site"
FT                  /*tag= bl

alignment_scores:
        Quality:    37.00          Length:        11
        Ratio:      3.700          Gaps:           0
Percent Similarity: 90.909   Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x T17521/rev

Align seg 1/1  to reverse of: T17521  from: 1  to: 24026

        2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
          |||||||||||||:::::::||  ||
    11314 AGATTGGCCATCAGAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17522

seq_documentation_block:
ID  T17522 standard; cDNA; 24026 BP.
AC  T17522;
DT  04-OCT-1996 (first entry)
DE  Mutated BRCA1 genomic sequence from PMA08.2.
KW  Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW  antibody production; germline alteration; probe; lesion neoplasia; human;
KW  gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  exon            256..355
FT                  /*tag= a
FT                  /note= "exon 1"
FT  intron          356..1512
FT                  /*tag= b
FT                  /note= "intron 1"
FT  misc_feature    1295
FT                  /*tag= c
FT                  /note= "known polymorphic site"
FT  exon            1513..1611
FT                  /*tag= d
FT                  /note= "exon 2"
FT  intron          1612..2206
FT                  /*tag= e
FT                  /note= "intron 2"
FT  misc_feature    1925..1937
FT                  /*tag= f
FT                  /note= "indefinite interval within intron 2"
FT  misc_feature    2141
FT                  /*tag= g
FT                  /note= "known polymorphic site"
FT  exon            2207..2260
FT                  /*tag= h
FT                  /note= "exon 3"
FT  intron          2261..2677
FT                  /*tag= i
FT                  /note= "intron 3"
FT  misc_feature    2569..2581
FT                  /*tag= j
FT                  /note= "indefinite interval within intron 3"
FT  exon            2678..2788
FT                  /*tag= k
FT                  /note= "exon 4"
FT  misc_feature    2725
FT                  /*tag= l
FT                  /note= "known polymorphic site"
FT  intron          2789..3328
FT                  /*tag= m
FT                  /note= "intron 4"
FT  misc_feature    3063..3075
FT                  /*tag= n
FT                  /note= "indefinite interval within intron 4"
FT  exon            3329..3406
FT                  /*tag= o
FT                  /note= "exon 5"
FT  intron          3407..3813
FT                  /*tag= p
FT                  /note= "intron 5"
FT  misc_feature    3598..3610
FT                  /*tag= q
FT                  /note= "indefinite interval within intron 5"
FT  misc_feature    3653
FT                  /*tag= r
FT                  /note= "known polymorphic site"
FT  exon            3814..3902
FT                  /*tag= s
FT                  /note= "exon 6"
FT  intron          3903..4224
FT                  /*tag= t
FT                  /note= "intron 6"
FT  misc_feature    4076..4088
FT                  /*tag= u
FT                  /note= "indefinite interval within intron 6"
FT  exon            4225..4364
FT                  /*tag= v
FT                  /note= "exon 7"
FT  intron          4365..6571
FT                  /*tag= w
FT                  /note= "intron 7"
FT  misc_feature    4391..4392
FT                  /*tag= x
FT                  /note= "known polymorphic site"
FT  misc_feature    4602..4614
FT                  /*tag= y
FT                  /note= "indefinite interval within intron 7"
FT  misc_feature    6538
FT                  /*tag= z
FT                  /note= "known polymorphic site"
FT  exon            6572..6677
FT                  /*tag= aa
FT                  /note= "exon 8"
FT  intron          6678..9163
FT                  /*tag= ab
FT                  /note= "intron 8"
FT  mutation        6823
FT                  /*tag= ac
FT                  /note= "A to T mutation at known polymorphic site"
FT  misc_feature    9106
FT                  /*tag= ad
FT                  /note= "known polymorphic site"
FT  exon            9164..9209
FT                  /*tag= ae
FT                  /note= "exon 9"
FT  misc_feature    9207
FT                  /*tag= af
FT                  /note= "known polymorphic site"
FT  intron          9210..10530
FT                  /*tag= ag
FT                  /note= "intron 9"
FT  misc_feature    9376
FT                  /*tag= ah
FT                  /note= "known polymorphic site"
FT  exon            10531..10607
FT                  /*tag= ai
```

```
FT                   /note= "exon 10"
FT   intron          10608..11597
FT                   /*tag= aj
FT                   /note= "intron 10"
FT   misc_feature    11384..11396
FT                   /*tag= ak
FT                   /note= "indefinite interval within intron 10"
FT   exon            11598..15023
FT                   /*tag= al
FT                   /note= "exon 11"
FT   misc_feature    11908
FT                   /*tag= am
FT                   /note= "known polymorphic site"
FT   misc_feature    11994
FT                   /*tag= an
FT                   /note= "known polymorphic site"
FT   misc_feature    12952
FT                   /*tag= ao
FT                   /note= "known polymorphic site"
FT   misc_feature    13004
FT                   /*tag= ap
FT                   /note= "known polymorphic site"
FT   misc_feature    13009
FT                   /*tag= aq
FT                   /note= "known polymorphic site"
FT   misc_feature    13048
FT                   /*tag= ar
FT                   /note= "known polymorphic site"
FT   misc_feature    13238
FT                   /*tag= as
FT                   /note= "known polymorphic site"
FT   misc_feature    13348
FT                   /*tag= at
FT                   /note= "known polymorphic site"
FT   misc_feature    13539
FT                   /*tag= au
FT                   /note= "known polymorphic site"
FT   misc_feature    13951
FT                   /*tag= av
FT                   /note= "known polymorphic site"
FT   misc_feature    14041
FT                   /*tag= aw
FT                   /note= "known polymorphic site"
FT   misc_feature    14046
FT                   /*tag= ax
FT                   /note= "known polymorphic site"
FT   misc_feature    14475
FT                   /*tag= ay
FT                   /note= "known polymorphic site"
FT   misc_feature    14874
FT                   /*tag= az
FT                   /note= "known polymorphic site"
FT   misc_feature    14891
FT                   /*tag= ba
FT                   /note= "known polymorphic site"
FT   misc_feature    14966
FT                   /*tag= bb
FT                   /note= "known polymorphic site"
FT   intron          15024..15424
FT                   /*tag= bc
FT                   /note= "intron 11"
FT   misc_feature    15284
FT                   /*tag= bd
FT                   /note= "known polymorphic site"
FT   exon            15425..15511
FT                   /*tag= be
FT                   /note= "exon 12"
FT   intron          15512..15952
FT                   /*tag= bf
FT                   /note= "intron 12"
FT   misc_feature    15647..15659
FT                   /*tag= bg
FT                   /note= "indefinite interval within intron 12"
FT   exon            15953..16126
FT                   /*tag= bh
FT                   /note= "exon 13"
FT   misc_feature    16077
FT                   /*tag= bi
FT                   /note= "known polymorphic site"
FT   intron          16127..16565
FT                   /*tag= bj
FT                   /note= "intron 13"
FT   misc_feature    16243
FT                   /*tag= bk
FT                   /note= "known polymorphic site"
FT   misc_feature    16370..16382
FT                   /*tag= bl
FT                   /note= "known polymorphic site"

alignment_scores:
                Quality:       37.00           Length:      11
                  Ratio:       3.700             Gaps:       0
     Percent Similarity:      90.909   Percent Identity:  63.636

alignment_block:
US-08-653-294-18 x T17522/rev

Align seg 1/1 to reverse of: T17522 from:   1  to: 24026

        2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
          |||||||||||||||::::::|||  |||
    11314 AGATTGGCCATCAGAAAACTGGTTCTGATGTAC 11282

seq_name: N_Geneseq_36:T17523

seq_documentation_block:
ID   T17523 standard; cDNA; 24026 BP.
AC   T17523;
DT   04-OCT-1996 (first entry)
DE   Mutated BRCA1 genomic sequence from PMA09.2.
KW   Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW   antibody production; germline alteration; probe; lesion neoplasia; human;
KW   gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            256..355
FT                   /*tag= a
FT                   /note= "exon 1"
FT   intron          356..1512
FT                   /*tag= b
FT                   /note= "intron 1"
FT   misc_feature    1295
FT                   /*tag= c
FT                   /note= "known polymorphic site"
FT   exon            1513..1611
FT                   /*tag= d
FT                   /note= "exon 2"
FT   intron          1612..2206
FT                   /*tag= e
FT                   /note= "intron 2"
FT   misc_feature    1925..1937
FT                   /*tag= f
FT                   /note= "indefinite interval within intron 2"
FT   misc_feature    2141
FT                   /*tag= g
FT                   /note= "known polymorphic site"
FT   exon            2207..2260
FT                   /*tag= h
FT                   /note= "exon 3"
FT   intron          2261..2677
FT                   /*tag= i
FT                   /note= "intron 3"
FT   misc_feature    2569..2581
FT                   /*tag= j
FT                   /note= "indefinite interval within intron 3"
FT   exon            2678..2788
FT                   /*tag= k
```

```
FT   misc_feature    2725
FT                   /note= "exon 4"
FT                   /*tag= l
FT   intron          2789..3328
FT                   /note= "known polymorphic site"
FT                   /*tag= m
FT   misc_feature    3063..3075
FT                   /note= "intron 4"
FT                   /*tag= n
FT   exon            3329..3406
FT                   /note= "indefinite interval within intron 4"
FT                   /*tag= o
FT   intron          3407..3813
FT                   /note= "exon 5"
FT                   /*tag= p
FT   misc_feature    3598..3610
FT                   /note= "intron 5"
FT                   /*tag= q
FT   misc_feature    3653..
FT                   /note= "indefinite interval within intron 5"
FT                   /*tag= r
FT   exon            3814..3902
FT                   /note= "known polymorphic site"
FT                   /*tag= s
FT   intron          3903..4224
FT                   /note= "exon 6"
FT                   /*tag= t
FT   misc_feature    4076..4088
FT                   /note= "intron 6"
FT                   /*tag= u
FT   exon            4225..4364
FT                   /note= "indefinite interval within intron 6"
FT                   /*tag= v
FT   intron          4365..6571
FT                   /note= "exon 7"
FT                   /*tag= w
FT   misc_feature    4391..4392
FT                   /note= "intron 7"
FT                   /*tag= x
FT   misc_feature    4602..4614
FT                   /note= "known polymorphic site"
FT                   /*tag= y
FT   misc_feature    6538
FT                   /note= "indefinite interval within intron 7"
FT                   /*tag= z
FT   exon            6572..6677
FT                   /note= "known polymorphic site"
FT                   /*tag= aa
FT   intron          6678..9163
FT                   /note= "exon 8"
FT                   /*tag= ab
FT   misc_feature    9164..9209
FT                   /note= "intron 8"
FT                   /*tag= ac
FT   misc_feature    9207
FT                   /note= "known polymorphic site"
FT                   /*tag= ad
FT   exon            9210..10530
FT                   /note= "exon 9"
FT                   /*tag= ae
FT   misc_feature    9376
FT                   /note= "known polymorphic site"
FT                   /*tag= af
FT   intron          9376
FT                   /note= "intron 9"
FT                   /*tag= ag
FT   misc_feature    9376
FT                   /note= "known polymorphic site"
FT                   /*tag= ah
FT   mutation        9376
FT                   /note= "T to C mutation at known polymorphic site"

FT   exon            10531..10607
FT                   /note= "exon 10"
FT                   /*tag= ai
FT   intron          10608..11597
FT                   /note= "intron 10"
FT                   /*tag= aj
FT   misc_feature    11384..11396
FT                   /note= "indefinite interval within intron 10"
FT                   /*tag= ak
FT   exon            11598..15023
FT                   /note= "exon 11"
FT                   /*tag= al
FT   misc_feature    11908
FT                   /note= "known polymorphic site"
FT                   /*tag= am
FT   misc_feature    11994
FT                   /note= "known polymorphic site"
FT                   /*tag= an
FT   misc_feature    12952
FT                   /note= "known polymorphic site"
FT                   /*tag= ao
FT   misc_feature    13004
FT                   /note= "known polymorphic site"
FT                   /*tag= ap
FT   misc_feature    13009
FT                   /note= "known polymorphic site"
FT                   /*tag= aq
FT   misc_feature    13048
FT                   /note= "known polymorphic site"
FT                   /*tag= ar
FT   misc_feature    13238
FT                   /note= "known polymorphic site"
FT                   /*tag= as
FT   misc_feature    13448
FT                   /note= "known polymorphic site"
FT                   /*tag= at
FT   misc_feature    13539
FT                   /note= "known polymorphic site"
FT                   /*tag= au
FT   misc_feature    13951
FT                   /note= "known polymorphic site"
FT                   /*tag= av
FT   misc_feature    14041
FT                   /note= "known polymorphic site"
FT                   /*tag= aw
FT   misc_feature    14046
FT                   /note= "known polymorphic site"
FT                   /*tag= ax
FT   misc_feature    14475
FT                   /note= "known polymorphic site"
FT                   /*tag= ay
FT   misc_feature    14874
FT                   /note= "known polymorphic site"
FT                   /*tag= az
FT   misc_feature    14891
FT                   /note= "known polymorphic site"
FT                   /*tag= ba
FT   misc_feature    14966
FT                   /note= "known polymorphic site"
FT                   /*tag= bb
FT   intron          15024..15424
FT                   /note= "intron 11"
FT                   /*tag= bc
FT   misc_feature    15284
FT                   /note= "known polymorphic site"
FT                   /*tag= bd
FT   exon            15425..15511
FT                   /note= "exon 12"
FT                   /*tag= be
FT   intron          15512..15952
FT                   /note= "intron 12"
FT                   /*tag= bf
FT   misc_feature    15647..15659
```

```
FT                  /*tag=  bg
FT                  /note=  "indefinite interval within intron 12"
FT       exon       15953..16126
FT                  /*tag=  bh
FT                  /note=  "exon 13"
FT       misc_feature  16077
FT                  /*tag=  bi
FT                  /note=  "known polymorphic site"
FT       intron     16127..16565
FT                  /*tag=  bj
FT                  /note=  "intron 13"
FT       misc_feature  16243
FT                  /*tag=  bk
```

```
alignment_scores:
            Quality:    37.00        Length:       11
            Ratio:      3.700        Gaps:          0
Percent Similarity:  90.909   Percent Identity:  63.636

alignment_block:
US-08-653-294-18 x T17523/rev

Align seg 1/1  to reverse of: T17523   from: 1   to: 24026

         2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
           ||||||||||||||||::::::|||  |||
     11314 AGATTGGCCATCAGAAACTGGTTCTGATGTAC 11282
```

OM of: US-08-653-294-18 to: EST:*   out_format : pfs
Date: Feb 8, 2000 4:03 AM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -DEV=xlp
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn1_1/USPTO_spool/US0865329A/runat_04022000_160700_15770/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -YGAPOP=10.000 -QGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US0653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-18
Query length: 12
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 8553.360000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_gss9:AQ137850	+	44.00	148.64	16.32	461	! AQ137850 HS_3058_B2_C09_MF CIT
gb_gss6:AQ869567	+	42.00	141.78	39.36	438	! AQ869567 nbeb0035D05f CUGI Rice
gb_gss6:AQ862871	+	42.00	140.92	43.94	483	! AQ862871 nbeb0019L13f CUGI Rice
gb_est34:AV120004	+	40.00	139.07	55.68	259	! AV120004 Mus musculus
gb_est36:AV182651	+	40.00	137.78	65.69	300	! AV182651 Yuji Kohara u
gb_gss0:AQ864125	+	40.00	135.02	93.62	411	! AQ864125 nbeb0022F09f CUGI Rice
gb_gss13:AQ460667	+	40.00	134.37	101.86	443	! AQ460667 HS_5135_B2_B07_SP6E RF
gb_gss6:AQ857430	+	40.00	134.03	106.26	460	! AQ857430 nbeb0005D22f CUGI Rice
gb_est24:AI235399	+	40.00	133.27	117.24	502	! AI235399 EST231961 Normalized r
gb_est27:AI407553	+	40.00	133.11	119.61	511	! AI407553 EST235843 Normalized r
gb_gss15:AQ661929	+	40.00	132.98	121.72	519	! AQ661929 HS_2130_B2_H03_MR CIT
gb_gss7:AQ917001	+	40.00	132.93	122.51	522	! AQ917001 nbeb0067L07f CUGI Rice
gb_gss6:AQ857059	+	40.00	129.17	198.33	801	! AQ857059 nbeb0004H21f CUGI Rice
gb_gss11:AQ279730	+	39.00	133.90	108.16	308	! AQ279730 CITBI-E1-2513B14.TR CI
gb_gss19:AA748739	+	39.00	133.87	108.55	309	! AA748739 ny06e12.s1 NCI_CGAP_GC
gb_est19:AA748739	+	39.00	132.50	129.31	361	! AA389896 mv30f04.r1 NCI_CGAP_Lu
gb_est44:AW214737	+	39.00	131.83	141.06	390	! AW214737 uo99c07.y1 NCI_CGAP_Lu
gb_est9:AA119739	−	39.00	130.79	161.15	439	! AA119739 mn14d02.r1 Beddington
gb_est14:AA414705	+	39.00	130.77	161.56	440	! AA414705 vc69c04.s1 Knowles Sol
gb_est18:AA692200	+	39.00	130.23	173.17	468	! AA692200 vt19h09.r1 Barstead mo
gb_gss6:AQ863678	+	39.00	130.17	174.42	471	! AQ863678 nbeb0021F21f CUGI Rice
gb_est20:AA839058	+	39.00	129.90	180.68	486	! AA839058 vw47a11.r1 Soares mous
gb_est8:AA012276	+	39.00	129.88	181.10	487	! AA012276 TgESTzz16c11.r1 TgME49
gb_gss13:AQ438867	+	39.00	129.79	183.62	492	! AQ438867 HS_5077_B2_H02_T7A RPC
gb_est10:AA168371	−	39.00	129.77	183.62	493	! AA168371 mf28a04.r1 Soares mous
gb_est5:H72837	−	39.00	129.58	188.23	504	! H72837 ys06c01.s1 Soares fetal
gb_est12:AA289553	+	39.00	129.56	188.65	505	! AA289553 vb17b05.r1 Soares mous
gb_est11:AA239865	+	39.00	129.52	189.49	507	! AA239865 mx80f12.r1 Soares mous
gb_est9:AA103036	−	39.00	129.32	194.55	519	! AA103036 mo21d11.r1 Life Tech m
gb_est36:AI891896	+	39.00	129.29	195.39	521	! AI891896 ul60c04.x1 Sugano mous
gb_est10:AA174998	−	39.00	128.21	224.31	589	! AA174998 ms88h04.r1 Soares mous
gb_est20:AA869826	−	39.00	128.21	224.31	589	! AA869826 vq16c07.r1 Barstead st
gb_est14:AQ569045	+	39.00	128.11	227.31	596	! AA569045 HS_5344_A1_F08_T7A RPC
gb_est10:AA137602	−	39.00	127.97	231.17	605	! AA137602 mq99a02.r1 Soares mous
gb_est7:AA638393	+	39.00	127.72	238.92	623	! AA638393 vl99d11.r1 Soares mous
gb_est16:AA571878	−	39.00	127.55	244.11	635	! AA571878 vm04e12.r1 Knowles Sol
gb_gss15:AQ655284	+	39.00	127.54	244.54	636	! AQ655284 Sheared DNA-27K23.TF.S
gb_gss15:AQ649156	+	39.00	127.21	254.95	660	! AQ649156 Sheared DNA-5K21.TF Sh
gb_gss3:B08474	+	39.00	122.50	466.40	1129	! B08474 T12N24-T7 TAMU Arabidop
gb_est10:AI171093	+	38.00	137.43	68.92	136	! AI171093 ms50c07.r1 Life Tech m
gb_est39:AW124405	−	38.00	133.52	113.56	212	! AW124405 UI-M-BH2.1-ape-d-08-0
gb_est34:AV135435	+	38.00	130.53	166.57	298	! AV135435 AV135435 Mus musculus

gb_est33:AV117866   −   38.00 130.33 170.98   305 ! AV117866 AV117866 Mus muscul
gb_est23:AU022889   −   38.00 128.68 211.20   368 ! AU022889 AU022889 Mouse unfe
gb_gss12:AQ366741   +   38.00 127.60 242.44   416 ! AQ366741 HS_5038_B2_D09_SP6E

seq_name: gb_gss9:AQ137850

seq_documentation_block:
LOCUS   AQ137850   461 bp   DNA   GSS   24-SEP-1998
DEFINITION HS_3058_B2_C09_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3058 Col=18 Row=F, genomic survey
            sequence.
ACCESSION   AQ137850
VERSION   AQ137850.1 GI:3528503
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 461)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3058 row: F column: 18
            Class: BAC ends
            High quality sequence stop: 461.
FEATURES   Location/Qualifiers
     source   1..461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="plate=3058 Col=18 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT   132 a   125 c   102 g   102 t
ORIGIN

alignment_scores:
        Quality: 44.00   Length: 12
        Ratio: 4.000   Gaps: 0
        Percent Similarity: 91.667   Percent Identity: 75.000

alignment_block:
US-08-653-294-18 x AQ137850

Align seg 1/1 to: AQ137850   from: 1 to: 461

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      ::|||::|||:::::::|::|:|::|||| |||||
   18 CATCGACTCACTATAAGGCGAATTCTCGGCCAGGTAC 53

seq_name: gb_gss6:AQ869567

seq_documentation_block:
LOCUS   AQ869567   438 bp   DNA   GSS   03-NOV-1999
DEFINITION nbeb0035D05f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0035D05f, genomic survey sequence.
ACCESSION   AQ869567
VERSION   AQ869567.1 GI:6220018
KEYWORDS   GSS.

LOCUS   AQ862871   483 bp   DNA   GSS   03-NOV-1999
DEFINITION nbeb0019L13f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0019L13f, genomic survey sequence.
ACCESSION   AQ862871
VERSION   AQ862871.1 GI:6213328
KEYWORDS   GSS.
SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 483)
AUTHORS   Wing,R.A. and Dean.R.A.
TITLE   A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL   Unpublished (1998)
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 40
            High quality sequence stop: 411.
FEATURES   Location/Qualifiers
     source 1..483
             /organism="Oryza sativa"
             /strain="Japonica"
             /cultivar="Nipponbare"
             /db_xref="taxon:4530"
             /clone_lib="CUGI Rice BAC Library (EcoRI)"
             /clone="nbeb0019L13f"
             /tissue_type="Leaf"
             /lab_host="E. coli DH10B"
             /note="vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
             Rice is the most important food crop in the world. Half of
             the world population, especially those inhabiting highly
             populated areas of the humid tropics and subtropics, rely
             on rice as their primary source of carbohydrate.
             Monocotyledonous rice is a diploid plant (2n=24) with a
             haploid genome equivalent of 431 Mbp (Arumuganathan and
             Earle, 1991). The relatively small genome of rice, three
             times larger than that of Arabidopsis, makes it suitable
             for genomic studies. In order to facilitate positional
             cloning, physical mapping and genome sequencing of rice,
             we have constructed a BAC library from Oryza sativa,
             Nipponbare variety using EcoRI as the cloning enzyme. The
             library contains 55,296 clones with an average insert size
             of 121 Kb providing approximatley 15 haploid genome
             equivalents. The deep coverage allows the isolation a
             particular sequence with a probability of 99.9 %. Three
             high density filters, each containing 18,432 clones
             (doubly spotted), represent the whole library for colony
             screening and can be requested from the Clemson University
             BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT   132 a   96 c   101 g   151 t   3 others
ORIGIN

alignment_scores:
       Quality: 42.00   Length: 11
       Ratio: 4.200   Gaps: 0
Percent Similarity: 90.909   Percent Identity: 81.818

alignment_block:
US-08-653-294-18 x AQ862871

Align seq 1/1 to: AQ862871 from: 1 to: 483

2 ArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
  ||||:|||||||||||||||||||| |||

SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 438)
AUTHORS   Wing,R.A. and Dean.R.A.
TITLE   A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL   Unpublished (1998)
COMMENT   On Mar 23, 1999 this sequence version replaced gi:334665.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 24
            High quality sequence stop: 397.
FEATURES   Location/Qualifiers
     source 1..438
             /organism="Oryza sativa"
             /strain="Japonica"
             /cultivar="Nipponbare"
             /db_xref="taxon:4530"
             /clone_lib="CUGI Rice BAC Library (EcoRI)"
             /clone="nbeb0035D05f"
             /tissue_type="leaf"
             /lab_host="E. coli DH10B"
             /note="vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
             Rice is the most important food crop in the world. Half of
             the world population, especially those inhabiting highly
             populated areas of the humid tropics and subtropics, rely
             on rice as their primary source of carbohydrate.
             Monocotyledonous rice is a diploid plant (2n=24) with a
             haploid genome equivalent of 431 Mbp (Arumuganathan and
             Earle, 1991). The relatively small genome of rice, three
             times larger than that of Arabidopsis, makes it suitable
             for genomic studies. In order to facilitate positional
             cloning, physical mapping and genome sequencing of rice,
             we have constructed a BAC library from Oryza sativa,
             Nipponbare variety using EcoRI as the cloning enzyme. The
             library contains 55,296 clones with an average insert size
             of 121 Kb providing approximatley 15 haploid genome
             equivalents. The deep coverage allows the isolation a
             particular sequence with a probability of 99.9 %. Three
             high density filters, each containing 18,432 clones
             (doubly spotted), represent the whole library for colony
             screening and can be requested from the Clemson University
             BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT   121 a   86 c   65 g   164 t   2 others
ORIGIN

alignment_scores:
       Quality: 42.00   Length: 10
       Ratio: 4.200   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-18 x AQ869567

Align seq 1/1 to: AQ869567 from: 1 to: 438

3 LeuAlaIleArgArgIleLeuLeuArgTyr 12
  |||:||||||||||||||||||||||
15 CTCACTATACGGCGAATTCTATTAAGATAT 44

seq_name: gb_gss6:AQ862871

seq_documentation_block:

```
                                        22 AGACTCTCTATACGGCGAATTCTTTTAGCCTAT 54

seq_name: gb_est34:AV120004

seq_documentation_block:
LOCUS       AV120004      259 bp      mRNA        EST      30-JUN-1999
DEFINITION  AV120004 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA
            clone 2610307A11, mRNA sequence.
ACCESSION   AV120004
VERSION     AV120004.1  GI:5302155
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 259)
AUTHORS     Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,T., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188908.
            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES    Location/Qualifiers
     source 1..259
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="2610307A11"
            /clone_lib="Mus musculus C57BL/6J 10-day embryo"
            /sex="mixed"
            /dev_stage="10-day embryo"
BASE COUNT  70 a    57 c    48 g    84 t
ORIGIN

alignment_scores:
        Quality:      40.00            Length:    11
        Ratio:         3.636           Gaps:       0
Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
US-08-653-294-18 x AV120004/rev    ..

Align seg 1/1  to reverse of: AV120004   from: 1   to: 259

    2 ArgLeuAlaIleArgArgLeuLeuArgTyr 12
      :::|||||::::|||||||||||::::|||
  241 AAGTTAGCCCTAAAAGGATCCTATTAAAATAT 209

seq_name: gb_est36:AV182651

seq_documentation_block:
LOCUS       AV182651      300 bp      mRNA        EST      21-JUL-1999
DEFINITION  AV182651 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk64g7 3', mRNA sequence.
ACCESSION   AV182651
VERSION     AV182651.1  GI:5562552
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 300)
AUTHORS     Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
TITLE       Expressed genes in C.elegans
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189500.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES    Location/Qualifiers
     source 1..300
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk64g7"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT  105 a    68 c    54 g    71 t
ORIGIN

alignment_scores:
        Quality:      40.00            Length:    12
        Ratio:         4.000           Gaps:       0
Percent Similarity: 83.333   Percent Identity: 58.333

alignment_block:
US-08-653-294-18 x AV182651

Align seg 1/1  to: AV182651   from: 1   to: 300

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      |||||||||(      |||||::::|||:::|||
   85 TACCGACTACTATCGCGTAGAGTATTAATACAATAT 120

seq_name: gb_gss6:AQ864125

seq_documentation_block:
LOCUS       AQ864125      411 bp      DNA                 GSS      03-NOV-1999
DEFINITION  nbeb0022F09f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0022F09f, genomic survey sequence.
ACCESSION   AQ864125
VERSION     AQ864125.1  GI:6214582
KEYWORDS    GSS.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 411)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
```

```
                 Tel: 864 656 7288
                 Fax: 864 656 4293
                 Email: rwing@clemson.edu
                 Seq primer: TAATACGACTCACTATAGGG
                 Class: BAC ends
                 High quality sequence start: 26
                 High quality sequence stop: 385.
FEATURES             Location/Qualifiers
     source          1 ..411
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0022F09f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      153 a    54 c    47 g    156 t     1 others
ORIGIN

alignment_scores:
     Quality:       40.00                Length:       10
     Ratio:          4.000               Gaps:          0
Percent Similarity: 100.000       Percent Identity:  90.000

alignment_block:
US-08-653-294-18 x AQ460667

Align seg 1/1  to: AQ864125   from: 1  to: 411

    2 ArgLeuAlaIleArgArgIleLeuLeuArg 11
      |||||::||||||||||||||
   13 CGACTCACTATACGGCGAATTCTACTACGA 42

seq_name: gb_gss13:AQ460667

seq_documentation_block:
LOCUS       AQ460667     443 bp    DNA            GSS       23-APR-1999
DEFINITION  HS_5135_B2_B07_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=711 Col=14 Row=D, genomic survey sequence.
ACCESSION   AQ460667
VERSION     AQ460667.1  GI:4639219
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 711 row: D column: 14
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1 ..443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=711 Col=14 Row=D"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      134 a    95 c    83 g    130 t     1 others
ORIGIN

alignment_scores:
     Quality:       40.00                Length:       12
     Ratio:          3.333               Gaps:          0
Percent Similarity: 100.000       Percent Identity:  50.000

alignment_block:
US-08-653-294-18 x AQ460667

Align seg 1/1  to: AQ460667   from: 1  to: 443

    1 TyrArgLeuAlaIleArgArgIleLeuLeuArgTyr 12
      ||::::::::::||||||::||||||||||||::
  237 TACCAAATGACCCTCAGAAGACTTTGCTGCGGTTC 272

seq_name: gb_gss6:AQ857430

seq_documentation_block:
LOCUS       AQ857430     460 bp    DNA            GSS       03-NOV-1999
DEFINITION  nbeb0005D22f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0005D22f, genomic survey sequence.
ACCESSION   AQ857430
VERSION     AQ857430.1  GI:6207887
KEYWORDS    GSS.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
```

```
Class: BAC ends
High quality sequence start: 82
High quality sequence stop: 353.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0005D22f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa.
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT       152 a       61 c       77 g      170 t
ORIGIN

alignment_scores:
          Quality:       40.00             Length:        11
            Ratio:        3.636              Gaps:         0
Percent Similarity:     100.000    Percent Identity:  72.727

alignment_block:
US-08-653-294-18 x AQ857430    ..
Align seg 1/1   to: AQ857430   from: 1  to: 460

     2 ArgLeuAlaIleArgArgIleLeuArgTyr 12
       |||||::::||||||||||||::::|||
    12 CGACTCACTATACGGCGAATTCTTATAGAATAT 44

seq_name: gb_est24:AI235399

seq_documentation_block:
LOCUS       AI235399       502 bp     mRNA         EST       31-JAN-1999
DEFINITION  EST231961 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVCR36 3' end, mRNA sequence.
ACCESSION   AI235399
VERSION     AI235399.1  GI:3828905
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151609.
            Other_ESTs: TC63429
```

```
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES             Location/Qualifiers
     source          1..502
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2040962"
                     /db_xref="taxon:10118"
                     /clone="ROVCR36"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT       169 a       90 c       85 g      158 t
ORIGIN

alignment_scores:
          Quality:       40.00             Length:        12
            Ratio:        4.444              Gaps:         0
Percent Similarity:     75.000    Percent Identity:  75.000

alignment_block:
US-08-653-294-18 x AI235399/rev   ..
Align seg 1/1  to reverse of: AI235399  from: 1   to: 502

     1 TyrArgLeuAlaIleArgArgIleLeuArgTyr 12
       ||||||||||||          |||||||||||||
   499 TACAGACTTGCCTACAGCCTTATCTTATTGAGGTAT 464

seq_name: gb_est27:AI407553

seq_documentation_block:
LOCUS       AI407553       511 bp     mRNA         EST       09-FEB-1999
DEFINITION  EST235843 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVD217 3' end, mRNA sequence.
ACCESSION   AI407553
VERSION     AI407553.1  GI:4251057
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Feb 17, 1998 this sequence version replaced gi:2889619.
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..511
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="ROVD217"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT       151 a      108 c      108 g      144 t
ORIGIN
```

```
alignment_scores:
        Quality:        40.00                    Length:         12
        Ratio:          4.444                     Gaps:            0
Percent Similarity:     75.000    Percent Identity:    75.000

alignment_block:
US-08-653-294-18 x AI407553/rev     ..
Align seg 1/1  to reverse of: AI407553  from: 1  to: 511

   1 TyrArgLeuAlaIleAargArgIleLeuLeuArgTyr 12
     ||||||||||            |||||||||||||
 378 TACAGACTTGCCTACAGCCTACCCTTATCTTATTGAGGTAT 343

seq_name: gb_gss15:AQ661929

seq_documentation_block:
LOCUS       AQ661929        519 bp      DNA           GSS       23-JUN-1999
DEFINITION  HS_2130_B2_H03_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate-2130 Col=6 Row=P, genomic survey
            sequence.
ACCESSION   AQ661929
VERSION     AQ661929.1   GI:5169697
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 519)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3556096.
            Contact: Mahairas GG, Wallace JC, Hood L
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Plate: 2130  row: P  column: 6
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 519.
FEATURES            Location/Qualifiers
     source          1..519
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="plate-2130 Col=6 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-coli DH10B"
BASE COUNT    169 a    106 c    108 g    136 t
ORIGIN

alignment_scores:
        Quality:        40.00                    Length:         10
        Ratio:          4.000                     Gaps:            0
Percent Similarity:    100.000    Percent Identity:    70.000

alignment_block:
US-08-653-294-18 x AQ661929/rev    ..
```

```
Align seg 1/1  to reverse of: AQ661929  from: 1  to: 519

   1 TyrArgLeuAlaIleAargArgIleLeuLeu 10
     ||||||||||:::||||||||||::::::|||
 154 TATCGTTTGTCTATTAGAAGAGTGATGCTT 125

seq_name: gb_gss7:AQ917001

seq_documentation_block:
LOCUS       AQ917001        522 bp      DNA           GSS       02-DEC-1999
DEFINITION  nbeb0067L07f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0067L07f, genomic survey sequence.
ACCESSION   AQ917001
VERSION     AQ917001.1   GI:6513517
KEYWORDS    GSS.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 522)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 34
            High quality sequence stop: 383.
FEATURES            Location/Qualifiers
     source          1..522
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0067L07f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT    154 a    103 c    101 g    163 t    1 others
ORIGIN

alignment_scores:
        Quality:        40.00                    Length:         10
        Ratio:          4.000                     Gaps:            0
```

Percent Similarity: 100.000   Percent Identity: 90.000

```
alignment_block:
US-08-653-294-18 x AQ917001  ..
Align seg 1/1 to: AQ917001 from: 1 to: 522

    2 ArgLeuAlaIleArgArgIleLeuLeuArg 11
      |||||:::|||||||||||||||||||
   16 CGATTGACTATACGCCGAATTCTACTACGA 45
```

seq_name: gb_gss6:AQ857059

```
seq_documentation_block:
LOCUS       AQ857059      801 bp    DNA     GSS       03-NOV-1999
DEFINITION  nbeb0004H21f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0004H21f, genomic survey sequence.
ACCESSION   AQ857059
VERSION     AQ857059.1  GI:6207425
KEYWORDS    GSS.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 24
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1..801
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0004H21f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      259 a    158 c    162 g    220 t    2 others
ORIGIN
```

```
alignment_scores:
            Quality:   40.00        Length:    10
            Ratio:     4.000        Gaps:       0
Percent Similarity: 100.000   Percent Identity: 80.000

alignment_block:
US-08-653-294-18 x AQ857059  ..
Align seg 1/1 to: AQ857059 from: 1 to: 801

    3 LeuAlaIleArgArgIleLeuLeuArgTyr 12
      ||||::|||||||||||||||||::::||||
   15 CTCACTATACGGCGAATTCTTATGCGTTAC 44
```

seq_name: gb_gss11:AQ279730

```
seq_documentation_block:
LOCUS       AQ279730      308 bp    DNA     GSS       22-NOV-1998
DEFINITION  CITBI-E1-2513B14.TR CITBI-E1 Homo sapiens genomic clone 2513B14,
            genomic survey sequence.
ACCESSION   AQ279730
VERSION     AQ279730.1  GI:3905634
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CITBI-E1-2513B14.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..308
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2513B14"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT       75 a     76 c     55 g    102 t
ORIGIN

alignment_scores:
            Quality:   39.00        Length:    12
            Ratio:     3.900        Gaps:       0
Percent Similarity: 83.333   Percent Identity: 58.333

alignment_block:
US-08-653-294-18 x AQ279730  ..
Align seg 1/1 to: AQ279730 from: 1 to: 308
```

```
  1 TyrArgLeuAlaIleArgArgIleLeuArgTyr 12
    ||| |||||||::  |||::|||::|||||||
223 TACTCTCTGGCTTTGTACAGACTGCTAATGAGATAC 258
```

seq_name: gb_est19:AA748739

```
seq_documentation_block:
LOCUS       AA748739         309 bp    mRNA           EST       22-JAN-1998
DEFINITION  ny06e12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270990 3',
            mRNA sequence.
ACCESSION   AA748739
VERSION     AA748739.1  GI:2788697
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 309)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797612.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 674   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 256.
FEATURES          Location/Qualifiers
     source       1..309
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:1270990"
                  /clone_lib="NCI_CGAP_GCB1"
                  /tissue_type="germinal center B cell"
                  /lab_host="DH10B"
                  /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                  polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                  was prepared from human tonsillar cells enriched for
                  germinal center B cells by flow sorting (CD20+, IgD-),
                  provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                  (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                  primed with a Not I - oligo(dT) primer
                  [5'-TGTTACCAATCTGAAGTGGGAGCGGCGGCCGCCATTTTTTTTTTTTTTTTTT-
                  3']. Double-stranded cDNA was ligated to Eco RI adaptors
                  (Pharmacia), digested with Not I and cloned into the Not I
                  and Eco RI sites of the modified pT7T3 vector. Library
                  went through one round of normalization, and was
                  constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      95 a     52 c     55 g    107 t
ORIGIN
```

```
alignment_scores:
     Quality:     39.00              Length:     12
     Ratio:        4.333             Gaps:        0
Percent Similarity: 75.000    Percent Identity: 66.667
```

alignment_block:
US-08-653-294-18 x AA748739/rev

Align seg 1/1 to reverse of: AA748739   from: 1   to: 309

```
  1 TyrArgLeuAlaIleArgArgIleLeuArgTyr 12
    |||||| ||||||||::|||||| |||
278 TACCGCAGAATGATACGACGTGTACTGCTTCTCTAC 243
```

```
              GenCore version 4.5
     Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February  8, 2000, 01:29:40 ; Search time 122.56 seconds
                                           (without alignments)
                                           2.319 Million cell updates/sec

Title:             US-08-653-294-19
Perfect score:     54
Sequence:          1 YRLAIXRIALRY 12

Scoring table:     BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched:          188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:      188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing:   Minimum Match 0%
                   Listing first 45 summaries

Database :         A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES

Result       %    Query
No.   Score Match Length DB ID      Description
----  ----- ----- ------ -- ------  ---------------------------
  1     52   96.3     12   1 R95429  HLA-B2702 84-79-84
  2     52   96.3     12   1 W33798  Peptide B2702.84-7
  3     52   96.3     12   1 W33799  Immunomodulating d
  4     38   70.4     20   1 R92907  HLA-B2702 CTL.modu
  5     38   70.4     20   1 R95428  HLA-B2702 84-75-84
  6     38   70.4     20   1 W33778  Immunomodulating d
  7     33   61.1     20   1 R92909  HLA-B2702 CTL modu
  8     33   61.1     20   1 R92908  HLA-B2702 CTL modu
  9     33   61.1     20   1 W33791  Peptide B2702.84-7
 10     33   61.1     20   1 W33792  Peptide B2702.84-7
 11     32   59.3    111   1 R81449  Hepatitis GB virus
 12     31   57.4     20   1 R95430  HLA-B2702 84-75T7
 13     29   53.7      6   1 W47261  Immunomodulatory p
 14     29   53.7      6   1 W33780  Peptide #1 used in
 15     29   53.7     10   1 R41208  Peptide fragment o
 16     29   53.7     10   1 R83082  HLA-B2702 CTL modu
 17     29   53.7     10   1 R83094  HLA-B2702 CTL modu
 18     29   53.7     10   1 R95413  Alpha1-helix of HL
 19     29   53.7     10   1 R95425  HLA-B2702.75-84(D)
 20     29   53.7     10   1 W07512  T-cell modulating
 21     29   53.7     10   1 W07513  T-cell modulating
 22     29   53.7     10   1 W47265  Immunomodulatory p
 23     29   53.7     10   1 W47269  Immunomodulatory p
 24     29   53.7     10   1 W33784  Peptide B2702.75-8
 25     29   53.7     10   1 W33787  Peptide B2702.75-8
 26     29   53.7     15   1 R92912  HLA-B2702 CTL modu
 27     29   53.7     15   1 W33795  Peptide B2702.70-8
 28     29   53.7     18   1 R71429  Human MHC 1 alpha
 29     29   53.7     25   1 R41205  Peptide fragment o
 30     29   53.7     25   1 R48286  Peptide fragment o
 31     29   53.7     25   1 R83090  HLA-B2702 CTL modu
 32     29   53.7     25   1 R83093  HLA-B2702 CTL modulat
 33     29   53.7     25   1 R95416  HLA-B2702.60-84.C
 34     29   53.7     25   1 R95422  HLAB38.6084. Comps
 35     29   53.7     25   1 W33794
 36     29   53.7     25   1 Y06801
 37     29   53.7     25   1 W41965
 38     29   53.7     25   1 W41993
 39     29   53.7     25   1 W41994
 40     29   53.7     25   1 R03142
 41     29   53.7     25   1 R03144
 42     29   53.7     25   1 R12463
 43     29   53.7     25   1 W44156
 44     29   53.7     25   1 R76710
 45     29   53.7     25   1 R90926

Peptide B2702.60-8
Peptide Seq ID No:
Flea cysteine prot
Flea cysteine prot
Flea cysteine prot
sequence of HLA-Bw
Sequence of HLA-B5
HLA-Bw53 exon. HLA
Human neuronal nic
Recombinant cold-r
E. histolytica pyr
```

ALIGNMENTS

```
RESULT   1
R95429
ID   R95429 standard; peptide; 12 AA.
AC   R95429;
DT   12-NOV-1996  (first entry)
DE   HLA-B2702 84-79-84 palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation.
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-79-84 palindrome. These sequences can be used to isolate
CC   the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC   protein associated with T-cell activation in mammalian T-cells, and is
CC   also immunologically cross reactive with the heat shock protein Hsc70.
CC   p74 is found in a limited number of cell types, but is particularly
CC   expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC   cell with an amphoteric detergent, and then passed through an affinity
CC   column containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence   12 AA;

     Query Match             96.3%;   Score 52;  DB 1;  Length 12;
     Best Local Similarity   91.7%;   Pred. No. 0.00017;
     Matches   11;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy          1 YRLAIXRIALRY 12
              ||||| |||||
Db          1 YRLAIRRIALRY 12

RESULT   2
W33798
ID   W33798 standard; peptide; 12 AA.
AC   W33798;
```

DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-79/79-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B -
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or V; aa77 -
CC   D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a hydrophobic or
CC   small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 12 AA;

Query Match   96.3%; Score 52; DB 1; Length 12;
Best Local Similarity   91.7%; Pred. No. 0.00017;
Matches   11; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
                ||||| ||||||
Db   1 YRLAIRRIALRY 12

RESULT 3
W33799
ID   W33799 standard; peptide; 12 AA.
AC   W33799;
DT   19-JUN-1998 (first entry)
DE   W33799 standard; immunomodulating dimer peptide #3.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 17; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is

CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B -(R aa76-77L)(aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or
CC   V; aa77 - D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a
CC   hydrophobic or small amino acid; aa82 - R or L; aa83 - G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. They can also inhibit the
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence 12 AA;

Query Match   96.3%; Score 52; DB 1; Length 12;
Best Local Similarity   91.7%; Pred. No. 0.00017;
Matches   11; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
                ||||| ||||||
Db   1 YRLAIRRIALRY 12

RESULT 4
R92907
ID   R92907 standard; peptide; 20 AA.
AC   R92907;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match   70.4%; Score 38; DB 1; Length 20;
Best Local Similarity   55.0%; Pred. No. 0.16;
Matches   11; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

Qy   1 YRLAI------XRIALRY 12
                ||||| ||||||
Db   1 YRLAIRLNERRENLRIALRY 20

RESULT 5
R95428
ID R95428 standard; peptide; 20 AA.
AC R95428;
DT 12-NOV-1996 (first entry)
DE HLA-B2702 84-75-84 palindrome.
KW HLA: p74: alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW cytolysis; antigen presenting cell.
OS Synthetic.
PN WO9513288-A1.
PD 18-MAY-1995.
PF 10-NOV-1994; U12985.
PR 10-NOV-1993; US-150493.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI: 95-194027/25.
PT Compsns. comprising lymphoid surface membrane proteins - which may
PT inhibit cytolytic activity and differentiation of CTLs.
PS Example: Page 12: 29pp; English.
CC R95413, and R95415-R95431 represent palindromes and fragments of
CC human-leucocyte-associated antigens. This sequence represents the
CC HLA-B2702 84-75-84 palindrome. These sequences can be used to isolate
CC the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC protein associated with T-cell activation in mammalian T-cells, and is
CC also immunologically cross reactive with the heat shock protein Hsc70.
CC p74 is found in a limited number of cell types, but is particularly
CC expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC cell with an amphoteric detergent, and then passed through an affinity
CC column containing a covalently bound HLA-B2702 palindromic peptide.
CC Compositions comprising the extracellular fragment of p74 combined with
CC HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC compounds can be screened for their effect on the cytolytic activity of
CC T-cells, by combining them with the extracellular portion of p74 and
CC determining the amount of binding between the candidate compound and p74.
CC Modulation of CTL activity can be inhibited in a cellular composition
CC containing T-cells and antigen presenting cells (APCs), by adding to the
CC mix the extracellular portion of p74, in an amount sufficient to compete
CC with p74 for the binding of the p74 ligand.
SQ Sequence 20 AA;

Query Match 70.4%; Score 38; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. NO. 0.16;
Matches 11; Conservative 0; Mismatches 0; Indels 1; Gaps 8; Gaps 1;

Qy 1 YRLAI-------XRIALRY 12
        ||||| ||||||
Db 1 YRLAIRLNERRENLRIALRY 20

RESULT 6
W33778
ID W33778 standard; peptide; 20 AA.
AC W33778;
DT 19-JUN-1998 (first entry)
DE Immunomodulating dimer peptide #1.
KW Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW rejection.
OS Synthetic.
OS Homo sapiens.
PN WO9744331-A1.
PD 27-NOV-1997.
PF 22-MAY-1997; U08689.
PR 24-MAY-1996; US-653294.
PA (STRD) UNIV. LELAND STANFORD JUNIOR.
PI Beulow R, Clayberger C, Krensky AM;
DR WPI: 98-086550/08.
PT New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT alpha-1 domain, used for preventing rejection of transplants or

PT treating autoimmune diseases
PS Claim 16; Page 35; 41pp; English.
CC This sequence represents a specifically claimed immunomodulating
CC dimer peptide of the invention. A peptide-type compound or variant is
CC claimed which has immunomodulating activity, including the N-terminal
CC acylated and/or C-terminal amidated or esterified forms of up to 60
CC amino acids, where the peptide-type compound comprises the formula; A-B,
CC where A, B = (R aa76-77L) (aa79-84) or (aa79-84) (Laa77-76R); aa76 = E or
CC V: aa77 = D, S or N: aa79 = R or G: aa80 = I or N: aa81, aa84= a
CC hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC represents amino acid. The sequence in the brackets may optionally be
CC absent or truncated at any peptide type bond within the brackets. The
CC compounds comprise amino acid sequences related to a Class I HLA-B
CC alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC vitro. They can also be used in combination with antigenic peptides or
CC proteins of interest to activate CTLs. They can also inhibit the
CC proliferation of T cells in response to anti-CD3. The peptide can be
CC used for preventing rejection of transplants or for treating autoimmune
CC diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC The products can also be used for detection and diagnosis.
SQ Sequence 20 AA;

Query Match 70.4%; Score 38; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. No. 0.16;
Matches 11; Conservative 0; Mismatches 0; Indels 1; Gaps 8; Gaps 1;

Qy 1 YRLAI-------XRIALRY 12
        ||||| ||||||
Db 1 YRLAIRLNERRENLRIALRY 20

RESULT 7
R92909
ID R92909 standard; peptide; 20 AA.
AC R92909;
DT 16-MAY-1996 (first entry)
DE HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW immunosuppressant; graft versus host disorder; transplantation; therapy:
KW class I MHC; HLA-B2702.
OS Synthetic.
PN WO9526979-A1.
PD 12-OCT-1995.
PF 05-APR-1995; U04349.
PR 05-APR-1994; US-222851.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM, Parham P;
DR WPI: 95-35582/46.
PT Extension of acceptance period of transplants from MHC unmatched
PT donor hosts - using Class I B75-84 MHC antigen of the recipient
PT host
PS Example 15; Page 36; 80pp; English.
CC R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC class I major histocompatibility complex (MHC) antigens. This sequence
CC is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC I MHC HLA-B2702. These sequences can be used to extend the period of
CC acceptance by a recipient of a transplant from an MHC unmatched donor.
CC The peptides are administered to a patient in conjunction with a
CC subtherapeutic amount of an immunosuppressant. This is administered to
CC the patient for a limited period of time (compared to the lifetime
CC administration for current treatments). The peptides particularly
CC modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC of the patient.
SQ Sequence 20 AA;

Query Match 61.1%; Score 33; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 10; Conservative 0; Mismatches 0; Indels 2; Gaps 8; Gaps 1;

Qy 1 YRLAI-------XRIALRY 12

```
Db           |||||           |||||
           1 YRLAIRLNERRENLRTALRY 20

RESULT 8
R92908
ID   R92908 standard; peptide; 20 AA.
AC   R92908
DT   16-MAY-1996 (first entry)
DE   R92908 CTL modulating peptide (B2702.84-75(T)/75-84).
KW   HLA-B2702 CTL modulating peptide; CTL; major histocompatibility complex; MHC;
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO952979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI: 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host.
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match             61.1%;    Score 33;   DB 1;   Length 20;
Best Local Similarity   50.0%;    Pred. No. 1.6;
Matches  10;   Conservative  0;   Mismatches  0;   Indels  8;   Gaps  1;

Qy    1 YRLAI-------XRIALRY 12
          ||||      |||||||
Db    1 YRLATRLNERRENLRIALRY 20

RESULT 9
W33791
ID   W33791 standard; peptide; 20 AA.
AC   W33791
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75(T)/75-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI: 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match             61.1%;    Score 33;   DB 1;   Length 20;
Best Local Similarity   50.0%;    Pred. No. 1.6;
Matches  10;   Conservative  0;   Mismatches  0;   Indels  8;   Gaps  1;

Qy    1 YRLAI-------XRIALRY 12
          ||||      |||||||
Db    1 YRLATRLNERRENLRIALRY 20

RESULT 10
W33792
ID   W33792 standard; peptide; 20 AA.
AC   W33792
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-75/75-84T tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI: 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match             61.1%;    Score 33;   DB 1;   Length 20;
```

Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 10; Conservative 0; Mismatches 2; Indels 8; Gaps 1;

```
Qy    1 YRLAI-------XRIALRY 12
        |||||       | ||||
Db    1 YRLAIRLNERENLRTALRY 20
```

RESULT 11
R81449
ID R81449 standard; Protein; 111 AA.
AC R81449;
DT 02-JUL-1996 (first entry)
DE Hepatitis GB virus (HGBV) clone 50 protein prod.
KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone 50;
KW tamarin; infected plasma; lambda phage; cDNA library.
OS Hepatitis GB virus.
FH Key   Location/Qualifiers
FT misc_difference 46
FT   /note= "corresponding codon STOP codon"
FT misc_difference 67
FT   /note= "corresponding codon STOP codon"
FT misc_difference 111
FT   /note= "corresponding codon STOP codon"
PN WO9521922-A2.
PD 17-AUG-1995.
PF 14-FEB-1995; U02118.
PF 14-FEB-1995; US-196030.
PR 14-FEB-1994; US-242654.
PR 29-JUL-1994; US-283314.
PR 13-MAY-1994; US-344190.
PR 23-NOV-1994; US-344185.
PR 23-NOV-1994; US-344557.
PR 27-JAN-1995; US-344557.
PA (ABBO) ABBOTT LAB.
PI Buijk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI Simons JN;
DR N-PSDB; T00049.
DR WPI; 95-293123/38.
PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT for diagnosis and therapy of hepatitis GB virus
PS Example 5; Page 222: 661pp; English.
CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. The cDNA clone T00049,
CC which encodes the proteins R81447-50 and R82064/65 (the 6 possible
CC reading frames), was rescued from the lambda phage, searched
CC against a sequence database and found to be an unique HGBV
CC sequence. Reagents which comprise the HGBV DNA, or its protein
CC prods. can be used for the diagnosis, therapy or in a vaccine to
CC prevent HGBV infection.
SQ Sequence 111 AA;

Query Match   59.3%; Score 32; DB 1; Length 111;
Best Local Similarity   41.7%; Pred. No. 13;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy    1 YRLAIXRIALRY 12
        || |:|:|:|
Db   81 YLLRMSRVAIKY 92
```

RESULT 12
R95430
ID R95430 standard; peptide; 20 AA.
AC R95430;
DT 12-NOV-1996 (first entry)
DE HLA-B2702 84-75T/75-84T palindrome.
KW HLA; B2702 84-75T/75-84T palindrome; human-leucocyte-associated antigen; inhibitor;
KW T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW cytolysis; antigen presenting cell.
OS Synthetic.
PN WO9511288-A1.
PD 18-MAY-1995.
PF 10-NOV-1994; U12985.
PR 10-NOV-1993; US-150493.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI; 95-194027/25.
PT Compns. comprising lymphoid surface membrane proteins - which may
PT inhibit cytolytic activity and differentiation of CTLs.
PS Example; Page 12; 29pp; English.
CC R95413, and R95415-R95431 represent palindromes and fragments of
CC human-leucocyte-associated antigens. This sequence represents the
CC HLA-B2702 84-75T/75-84T palindrome. These sequences can be used to
CC isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface
CC membrane protein associated with T-cell activation in mammalian T-cells,
CC and is also immunologically cross reactive with the heat shock protein
CC Hsc70. p74 is found in a limited number of cell types, but is
CC particularly expressed on B and T cells. p74 can be isolated by lysis of
CC a suitable cell with an amphoteric detergent, and then passed through an
CC affinity column containing a covalently bound HLA-B2702 palindromic
CC peptide. Compositions comprising the extracellular fragment of p74
CC combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and
CC inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.
CC Candidate compounds can be screened for their effect on the cytolytic
CC activity of T-cells, by combining them with the extracellular portion of
CC p74 and determining the amount of binding between the candidate compound
CC and p74. Modulation of CTL activity can be inhibited in a cellular
CC composition containing T-cells and antigen presenting cells (APCs), by
CC adding to the mix the extracellular portion of p74, in an amount
CC sufficient to compete with p74 for the binding of the p74 ligand.
SQ Sequence 20 AA;

Query Match   57.4%; Score 31; DB 1; Length 20;
Best Local Similarity   52.6%; Pred. No. 3.8;
Matches 10; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

```
Qy    1 YRLAI-------XRIALR 11
        |||||       | ||||
Db    1 YRLAIRLNETRENLRIALR 19
```

RESULT 13
W47261
ID W47261 standard; peptide; 6 AA.
AC W47261;
DT 22-MAY-1998 (first entry)
DE Immunomodulatory peptide.
KW Immunomodulator; Class I HLA-B alpha-1 domain; inhibition;
KW transplant rejection; treatment; autoimmune disease.
OS Homo sapiens.
OS Synthetic.
FH Key   Location/Qualifiers
FT Misc_difference 1..6
FT   /note= "at least one of the amino acids is the
FT   D-isomer
PN WO9744052-A1.
PD 27-NOV-1997.
PF 23-APR-1997; U06705.
PR 22-MAY-1996; US-651650.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Clayberger C, Krensky AM;
DR WPI; 98-018220/02.
PT Novel immunomodulatory peptide-type compound - useful for inhibiting
PT transplant rejection
PS Claim 10; Page 36; 41pp; English.
CC The present sequence is an immunomodulatory peptide, which
CC comprises a Class I HLA-B alpha-1 domain sequence. It can be used
CC in a pharmaceutical composition together with a subtherapeutic dose
CC of an immunosuppressant, to extend the period of acceptance of a
CC transplant from a major histocompatibility complex (MHC) unmatched

CC   donor, i.e. to inhibit transplant rejection. It can also be used in
CC   the treatment of autoimmune diseases.
CC   Peptides using the D-form amino acids are more effective
CC   immunomodulators than their diastereomers or enantiomers.
SQ   Sequence   6 AA;

Query Match   53.7%; Score 29; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   7 RIALRY 12
         ||||||
Db   1 RIALRY   6

RESULT 14
W33780
ID   W33780 standard: peptide: 6 AA.
AC   W33780;
DT   19-JUN-1998 (first entry)
DE   Peptide #1 used in immunomodulating dimer peptide.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation: autoimmune disease: Class I HLA-B alpha-1 domain:
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-Al.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI: 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 15; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed peptide which forms part
CC   of the immunomodulating dimer peptides of the invention. A peptide-type
CC   compound or variant is claimed which has immunomodulating activity,
CC   including the N-terminal acylated and/or C-terminal amidated or
CC   esterified forms of up to 60 amino acids, where the peptide-type compound
CC   comprises the formula; A-B, where A, B -(R aa76-77L) (aa79-84) or
CC   (aa84-79) (Laa77-76R); aa76 - E or V; aa77 - D, S or N; aa79 - R or G;
CC   aa80 - I or N; aa81, aa84= a hydrophobic or small amino acid; aa82 = R
CC   or L; aa83 - G or R; and aa represents amino acid. The sequence in the
CC   brackets may optionally be absent or truncated at any peptide type bond
CC   within the brackets. The compounds comprise amino acid sequences related
CC   to a Class I HLA-B alphal domain (positions 79-84). They can be used to
CC   inhibit cytotoxic T-lymphocytes (CTL) from undesirably attacking cells in
CC   a host or in vitro. They can also be used in combination with antigenic
CC   peptides or proteins of interest to activate CTLs. They can also inhibit
CC   the proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence   6 AA;

Query Match   53.7%; Score 29; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.5e+05;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   7 RIALRY 12
         ||||||
Db   1 RIALRY   6

RESULT 15
R41208
ID   R41208 standard: peptide: 10 AA.
AC   R41208;

DT   15-MAR-1994 (first entry)
DE   Peptide fragment of Class I HLA peptide.
KW   Human leukocyte antigen; HLA; peptide; transplantation; neoplasia;
KW   parasitic disease; cytotoxic T lymphocyte; modulation.
OS   Synthetic.
PN   WO9317699-A.
PD   16-SEP-1993.
PF   25-FEB-1993; U01758.
PR   02-MAR-1992; US-844716.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger CA, Krensky AM;
DR   WPI: 93-303134/38.
PT   New peptide(s) based on Class I HLA antigen domains - used for
PT   modulating cytotoxic T-lymphocyte activity towards targets
PS   Claim 11; Page 54; 61pp; English.
CC   The peptide is used to modulate cytotoxic T-lymphocyte (CTL)
CC   activity, either by inhibition or stimulation. It can be used
CC   for inhibiting CTL toxicity in transplantations, for inducing CTL
CC   activity in parasitic diseases and neoplasia and in studies on viral
CC   infection. The peptide can also be used for identifying CTLs which
CC   bind to it and removing subsets of CTLs from a T-cell composition.
CC   This peptide sequence is more commonly found within larger peptide
CC   compounds of not more than 30 amino acids in length.
SQ   Sequence   10 AA;

Query Match   53.7%; Score 29; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   7 RIALRY 12
         ||||||
Db   5 RIALRY 10

Search completed: February 8, 2000, 01:29:40
Job time: 1752 sec

```
                    GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 7, 2000, 11:54:29 ; Search time 117.7 Seconds
                                         (without alignments)
                                         4.809 Million cell updates/sec

Title:           US-08-653-294-19
Perfect score:   54
Sequence:        1 YRLAIXRIALRY 12

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters:     142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08
                 Listing first 45 summaries

Database :    PIR_62:*
              1:  pir1:*
              2:  pir2:*
              3:  pir3:*
              4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	63.0	141	2	H71504	ribosomal protein
2	34	63.0	309	2	G70882	probable oxidoredu
3	33	61.1	142	2	C72053	L17 ribosomal prot
4	32	59.3	253	2	E75055	hypothetical prote
5	32	59.3	264	2	E71567	hypothetical prote
6	32	59.3	552	2	A64085	probable ATP-bindi
7	32	59.3	610	2	S77337	ABC-type transport
8	32	59.3	2242	2	A57541	pyrimidine synthes
9	31	57.4	190	2	F70410	hypothetical prote
10	31	57.4	349	2	G70542	probable bioB prot
11	31	57.4	356	2	S37356	spaS protein - Sal
12	31	57.4	469	2	C69460	conserved hypothet
13	31	57.4	578	2	B64012	hypothetical prote
14	31	57.4	1166	2	A49201	adenylate cyclase
15	31	57.4	1180	2	A47202	adenylate cyclase
16	31	57.4	1791	2	T02909	hypothetical prote
17	30	55.6	88	2	C70200	ribosomal protein
18	30	55.6	153	2	S08501	3-dehydroquinate d
19	30	55.6	211	2	H64961	probable membrane
20	30	55.6	283	2	JC6531	avermectin B 5-O-m
21	30	55.6	286	1	B62290	rhamnosyl transfer
22	30	55.6	313	2	JC5542	Na+/H+ antiporter
23	30	55.6	331	2	S09800	hypothetical prote
24	30	55.6	348	2	S29990	histocompatibility
25	30	55.6	376	2	S75438	hypothetical prote
26	30	55.6	377	2	A72350	pleiotropic regula
27	30	55.6	394	2	S20905	hypothetical prote
28	30	55.6	432	2	F70575	probable PurA - My
29	30	55.6	497	2	D65189	yigC protein - Esc
30	30	55.6	528	2	B64760	Propionate catabol
31	30	55.6	564	2	I64134	D-lactate dehydrog
32	30	55.6	690	2	S54211	ATM1 protein precu
33	30	55.6	754	2	S52564	hypothetical prote
34	30	55.6	770	1	WZBE30	gene 30 protein -
35	30	55.6	880	2	F71652	pyruvate,phosphate
36	30	55.6	1084	2	G71329	hypothetical prote
37	30	55.6	2347	1	TVHURS	kinase-related pro
38	30	55.6	3712	2	S18253	laminin alpha-1 ch
39	29	53.7	124	1	QQV2EL	F11 protein - vacc
40	29	53.7	133	2	H42513	J5L protein - vacc
41	29	53.7	133	2	S33096	J5L protein - vari
42	29	53.7	138	1	A69219	conserved hypothet
43	29	53.7	144	2	H72464	hypothetical prote
44	29	53.7	156	2	T12893	hypothetical prote
45	29	53.7	170	2	E69149	hypothetical prote

ALIGNMENTS

```
RESULT    1
H71504
ribosomal protein L17 - Chlamydia trachomatis
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 13-Aug-1999
C;Accession: H71504; I40747
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809
A;Accession: H71504
A;Molecule type: DNA
A;Residues: 1-141 <ARN>
A;cross-references: GB:AE001323; GB:AE001273; NID:g3328931; PIDN:AAC68107.1; PID:g332
A;Experimental source: serotype D, strain UW-3/Cx
R;Gu, L.; Wenman, W.M.; Remacha, M.; Meuser, R.; Coffin, J.; Kaul, R.
J. Bacteriol. 177, 2594-2601, 1995
A;Title: Chlamydia trachomatis RNA polymerase alpha subunit: sequence and structural
A;Reference number: I40743; MUID:95247702
A;Accession: I40747
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 20-115,'R',117-141 <GUL>
A;Cross-references: GB:L33834; NID:g620026; PIDN:AAA74990.1; PID:g620030
C;Genetics:
A;Gene: rl17
C;Superfamily: Escherichia coli ribosomal protein L17
C;Keywords: protein biosynthesis; ribosome


Query Match             63.0%;  Score 34;  DB 2;  Length 141;
Best Local Similarity   54.5%;  Pred. No. 3.7;
Matches    6;  Conservative   3;  Mismatches   2;  Indels    0;  Gaps    0;

Qy         2 RLAIXRIALRY 12
             ||: |: :||
Db        64 RLAVGRLMVRY 74


RESULT    2
G70882
probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 29-Sep-1999
C;Accession: G70882
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Nature 393, 537-544, 1998
A;Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Qua
.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: G70882
```

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-309 <COL>
A;Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA11591.1; PID:e129972
C;Genetics:
A;Gene: Rv2776c
C;Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred
F;9-211/Domain: cytochrome-b5 reductase homology <CBR>
F;240-297/Domain: ferredoxin [2Fe-2S] homology <FER>

Query Match   63.0%; Score 34; DB 2; Length 309;
Best Local Similarity 77.8%; Pred. No. 8;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 YRLAIXRIA 9
            || :|| |||
Db   65 YRIAIRRIA 73

RESULT   3
C72053
L17 ribosomal protein - Chlamydia pneumoniae (strain CWL029)
C;Species: Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999
C;Accession: C72053
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606
A;Accession: C72053
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-142 <ARN>
A;Cross-references: GB:AE001647; GB:AE001363; NID:g4376920; PIDN:AAD18764.1; PID:g437692
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: rl17
C;Superfamily: Escherichia coli ribosomal protein L17

Query Match   61.1%; Score 33; DB 2; Length 142;
Best Local Similarity 54.5%; Pred. No. 6.1;
Matches   6; Conservative   3; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIXRIALRY 12
            |:|| |: :||
Db   64 RIAIGRLMVRY 74

RESULT   4
E75055
hypothetical protein PAB1428 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: E75055
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: E75055
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <KAW>
A;Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50338.1; PID:e151623
C;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1428

Query Match   59.3%; Score 32; DB 2; Length 253;
Best Local Similarity 66.7%; Pred. No. 18;

Matches   8; Conservative   0; Mismatches   0; Indels   4; Gaps   0;

Qy   1 YRLAIXRIALRY 12
            | ||| ||| | ||
Db   141 YLLAIPRSALTY 152

RESULT   5
E71567
hypothetical protein ybbP - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 26-Aug-1999
C;Accession: E71567
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809
A;Accession: E71567
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-264 <ARN>
A;Cross-references: GB:AE001276; GB:AE001273; NID:g3328399; PID:g3328401
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: ybbP
C;Superfamily: hypothetical protein ybbP

Query Match   59.3%; Score 32; DB 2; Length 264;
Best Local Similarity 70.0%; Pred. No. 18;
Matches   7; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIXRIALR 11
            ||| :| ||
Db   83 RLALSRIRLR 92

RESULT   6
A64085
probable ATP-binding transport protein HI0664 - Haemophilus influenzae (strain Rd KW2
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 20-Sep-1999
C;Accession: A64085
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
, E.A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: A64085
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-552 <TIGR>
A;Cross-references: GB:U32749; GB:I42023; NID:g1573658; PIDN:AAC22321.1; PID:g1573663
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C;Keywords: ATP; P-loop
F;355-549/Domain: ATP-binding cassette homology <ABC>
F;372-379/Region: nucleotide-binding motif A (P-loop)

Query Match   59.3%; Score 32; DB 2; Length 552;
Best Local Similarity 70.0%; Pred. No. 38;
Matches   7; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   3 LAIXRIALRY 12
            ||| || |||
Db   70 LAVARGALRY 79

RESULT   7
S77337
ABC-type transport protein sll1725 - Synechocystis sp. (strain PCC 6803)

N:Alternate names: ABC transporter; protein sll1725
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Sep-1999
C:Accession: S77337
R:KaneKo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
s.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
A:Reference number: S74322; MUID:97061201
A:Accession: S77337
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-610 <KAN>
A:Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAA17440.1; PID:d101817
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; P-loop; transport protein
F:380-574/Domain: ATP-binding cassette homology <ABC>
F:397-404/Region: nucleotide-binding motif A (P-loop)

Query Match   59.3%; Score 32; DB 2; Length 610;
Best Local Similarity 70.0%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

Qy   2 RLAIXRIALR 11
          ||| | |:|
Db 508 RLAIARAAMR 517

RESULT 8
A57541
pyrimidine synthesis multifunctional protein CAD - spiny dogfish
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase (g
C:Species: Squalus acanthias (spiny dogfish)
C:Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 22-Jun-1999
C:Accession: A57541
R:Hong, J.; Salo, W.L.; Anderson, P.M.
J. Biol. Chem. 270, 14130-14139, 1995
A:Title: Nucleotide sequence and tissue-specific expression of the multifunctional prote
A:Reference number: A57541; MUID:95294021
A:Accession: A57541
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2242 <HON>
A:Cross-references: GB:U18868; NID:g951095; PIDN:AAA74569.1; PID:g951096
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; Ba
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
C:Keywords: hydrolase; ligase; transferase
F:4-149/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:4-354/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homolog
F:178-354/Domain: trpG homology <TRG>
F:398-1446/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homo
F:398-849/Domain: biotin carboxylase homology <BC1>
F:939-1385/Domain: biotin carboxylase homology <BC2>
F:1464-1808/Domain: Bacillus dihydroorotase homology <DHO>
F:1941-2239/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
F:252/Active site: Cys #status predicted

Query Match   59.3%; Score 32; DB 2; Length 2242;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy   1 YRLAIXRIALRY 12
         || |:|:|| |||
Db 2109 YLLTLYRVNLRY 2120

RESULT 9
F70410

hypothetical protein aq_1277 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
C:Accession: F70410
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:9819666
A:Accession: F70410
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <AQF>
A:Cross-references: GB:AE000732; NID:g2983704; PID:g2983715; GB:AE000657
C:Genetics:
A:Gene: aq_1277

Query Match   57.4%; Score 31; DB 2; Length 190;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy   1 YRLAIXRIALR 11
         | ||| :||:|
Db   43 HNLAIQKVALR 53

RESULT 10
G70542
probable bioB protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Jul-1999
C:Accession: G70542
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:9829587
A:Accession: G70542
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-349 <COL>
A:Cross-references: GB:295586; GB:AL123456; NID:g3261785; PIDN:CAB09080.1; PID:e31716
C:Genetics:
A:Gene: bioB
C:Superfamily: biotin synthetase

Query Match   57.4%; Score 31; DB 2; Length 349;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy   1 YRLAIXRIALRY 12
         ||| :|: ||:
Db 280 FRLALPRTMLRF 291

RESULT 11
S37356
spaS protein - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Sep-1999
C:Accession: S37356
R:Groisman, E.A.; Ochman, H.
EMBO J. 12, 3779-3787, 1993
A:Title: Cognate gene clusters govern invasion of host epithelial cells by Salmonella
A:Reference number: S37304; MUID:94008985
A:Accession: S37356
A:Status: nucleic acid sequence not shown; translation not shown

```
Qy      1 YRLAIXRIAL 10
          |:|| |:|||
Db    147 YKTAISRLAL 156
```

RESULT 14
A49201
adenylate cyclase (EC 4.6.1.1) type V, calcium-inhibitable - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A49201; A46187
R;Premont, R.T.; Jacobowitz, O.; Iyengar, R.
Endocrinology 131, 2774-2784, 1992
A;Title: Lowered responsiveness of the catalyst of adenylyl cyclase to stimulation by
A;Reference number: A49201; MUID:93076707
A;Accession: A49201
A;Molecule type: mRNA
A;Residues: 7-1161 <PRE>
A;Cross-references: GB:M96653; NID:g191726; PIDN:AAA37182.1; PID:g191727
A;Experimental source: S49 lymphoma cells
A;Note: sequence extracted from NCBI backbone (NCBIN:119384, NCBIP:119386)
R;Yoshimura, M.; Cooper, D.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 6716-6720, 1992
A;Title: Cloning and expression of a Ca(2+)-inhibitable adenylyl cyclase from NCB-20
A;Reference number: A46187; MUID:92357702
A;Accession: A46187
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-46,'K',48-75,'G',77-507,'GR',511-737,'v',739-881,'L',883-990,'v',992-11
A;Cross-references: GB:M93422; NID:g191690; PIDN:AAA37174.1; PID:g191691
A;Experimental source: NCB-20 cells
A;Note: sequence extracted from NCBI backbone (NCBIP:110233); the authors acknowledge
C;Superfamily: human adenylate cyclase; guanylate cyclase catalytic domain homology <GCC>
C;Keywords: phosphorus-oxygen lyase; transmembrane protein
F;319-554/Domain: guanylate cyclase catalytic domain homology <GCC>
F;925-1164/Domain: guanylate cyclase catalytic domain homology <GCC2>

Query Match   57.4%; Score 31; DB 2; Length 1166;
Best Local Similarity   55.6%;   Pred. No. 1.3e+02;
Matches   5; Conservative   3; Mismatches   1; Indels   1; Gaps   0;

```
Qy      4 AIXRIALRY 12
          |:|:|||:|
Db    886 AVGRVALKY 894
```

RESULT 15
A47202
adenylate cyclase (EC 4.6.1.1) type 6 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A47202; A45145
R;Premont, R.T.; Chen, J.; Ma, H.W.; Ponnapalli, M.; Iyengar, R.
Proc. Natl. Acad. Sci. U.S.A. 89, 9809-9813, 1992
A;Title: Two members of a widely expressed subfamily of hormone-stimulated adenylyl c
A;Reference number: A47202; MUID:93028552
A;Accession: A47202
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1180 <PRE>
A;Experimental source: liver, kidney
A;Note: sequence extracted from NCBI backbone (NCBIP:115851)
R;Krupinski, J.; Lehman, T.C.; Frankenfield, C.D.; Zwaagstra, J.C.; Watson, P.A.
J. Biol. Chem. 267, 24858-24862, 1992
A;Title: Molecular diversity in the adenylylcyclase family. Evidence for eight forms
A;Reference number: A45145; MUID:93077589
A;Accession: A45145
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 15-93,'K',95-463,'v',465-551,'G',553-803,'I',805-1180 <KRU>
A;Cross-references: GB:L01115; NID:g202712; PIDN:AAA40676.1; PID:g202713

A;Molecule type: DNA
A;Residues: 1-356 <GRO>
A;Cross-references: EMBL:X73525; NID:g404286; PID:g404293
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
C;Genetics:
A;Gene: spaS
C;Superfamily: flagellar biosynthetic protein flhB; flhB carboxyl-terminal homology
C;Keywords: transmembrane protein

Query Match   57.4%; Score 31; DB 2; Length 356;
Best Local Similarity   50.0%;   Pred. No. 40;
Matches   6; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

```
Qy      1 YRLAIXRIALRY 12
          |:|| :|:|
Db     70 YSLAVFGIGLKY 81
```

RESULT 12
C69460
conserved hypothetical protein AF1684 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: C69460
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A;Reference number: A69250; MUID:98049343
A;Accession: C69460
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-469 <KLE>
A;Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89565.1; PID:g264887
C;Superfamily: hypothetical protein MJ0966

Query Match   57.4%; Score 31; DB 2; Length 469;
Best Local Similarity   58.3%;   Pred. No. 53;
Matches   7; Conservative   1; Mismatches   4; Indels   0; Gaps   0;

```
Qy      1 YRLAIXRIALRY 12
          || :|||
Db     12 YRKGMKRIALVY 23
```

RESULT 13
B64012
hypothetical protein HI0698 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 10-Oct-1997
C;Accession: B64012
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; D.M.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: B64012
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-578 <TIGR>
A;Cross-references: GB:U32752; GB:L42023; NID:g1573692; PID:g1573700; TIGR:HI0698

Query Match   57.4%; Score 31; DB 2; Length 578;
Best Local Similarity   60.0%;   Pred. No. 65;
Matches   6; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

C:Superfamily: human adenylate cyclase; guanylate cyclase catalytic domain homology
C:Keywords: phosphorus-oxygen lyase
F:333-568/Domain: guanylate cyclase catalytic domain homology <GCC>
F:939-1178/Domain: guanylate cyclase catalytic domain homology <GCC2>

Query Match   57.4%; Score 31; DB 2; Length 1180;
  Best Local Similarity 55.6%; Pred. No. 1.3e+02;
  Matches   5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy   4 AIXRIALRY 12
           |: |:||:|
Db   900 AVGRVALKY 908

Search completed: February 7, 2000, 11:54:30
Job time: 24340 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February  8, 2000, 00:59:57 ; Search time 63.71 Seconds
                                          (without alignments)
                                          5.625 Million cell updates/sec

Title:          US-08-653-294-19
Perfect score:  54
Sequence:       1 YRLAIXRIALRY 12

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       82229 seqs, 29964866 residues

Total number of hits satisfying chosen parameters:    82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :     SwissProt_38:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                              SUMMARIES
            %
Result      Query
No.   Score Match Length DB   ID          Description
------ ----- ----- ------ --  ----------  -----------------------
  1    34   63.0   122   1    RL17_CHLTR  P47760 chlamydia t
  2    33   61.1   375   1    HRMA_PSESY  Q08370 pseudomonas
  3    32   59.3   552   1    Y664_HAEIN  Q57538 haemophilus
  4    32   59.3  2242   1    PYR1_SQUAC  Q91437 squalus aca
  5    31   57.4   349   1    BIOB_MYCTU  O06601 mycobacteri
  6    31   57.4   356   1    SPAS_SALTY  P40702 salmonella
  7    31   57.4   456   1    YD45_SCHPO  Q10299 schizosacch
  8    31   57.4   541   1    PRPR_SALTY  P74839 salmonella
  9    31   57.4   578   1    YTFM_HAEIN  P44038 haemophilus
 10    31   57.4   582   1    MNT_HUMAN   Q99583 homo sapien
 11    31   57.4   591   1    MNT_MOUSE   O08789 mus musculu
 12    31   57.4   833   1    YPD2_CAEEL  Q11182 caenorhabdi
 13    31   57.4  1165   1    CYA6_MOUSE  Q01341 mus musculu
 14    31   57.4  1166   1    CYA6_RAT    Q03343 rattus norv
 15    31   57.4  2569   1    LMA3_MOUSE  Q61789 mus musculu
 16    30   55.6    88   1    RS15_BORBU  O51744 borrelia bu
 17    30   55.6   153   1    3DHQ_EMENI  P05147 emericella
 18    30   55.6   331   1    UL38_HCMVA  P16779 human cytom
 19    30   55.6   348   1    HLAF_MACMU  P33617 macaca mula
 20    30   55.6   432   1    PURA_MYCTU  O08381 mycobacteri
 21    30   55.6   497   1    YIGC_ECOLI  P26615 escherichia
 22    30   55.6   528   1    PRPR_ECOLI  P77743 escherichia
 23    30   55.6   564   1    LDHD_HAEIN  P45295 haemophilus
 24    30   55.6   690   1    ATM1_YEAST  P40416 saccharomyc
 25    30   55.6   770   1    PRTP_VZVD   P09284 varicella-z
 26    30   55.6  2347   1    KROS_HUMAN  P08922 homo sapien
 27    30   55.6  3712   1    LMA_DROME   Q00174 drosophila
 28    29   53.7   124   1    VJ05_VACCV  P07618 vaccinia vi
 29    29   53.7   133   1    VJ05_VACCC  P21083 vaccinia vi
 29    29   53.7   133   1    VJ05_VARV   P33055 variola vir
 30    29   53.7   185   1    PTH_RICPR   Q92cv4 rickettsia
 31    29   53.7   203   1    NH1O_YEAST  Q03435 saccharomyc
 32    29   53.7   281   1    STRF_STRGR  P09397 streptomyce
 33    29   53.7   337   1    CMST_HUMAN  P78382 homo sapien
 34    29   53.7
```

```
 35  29  53.7  359  1    1B01_PANTR      P13750 pan troglod
 36  29  53.7  362  1    1B01_GORGO      P30379 gorilla gor
 37  29  53.7  362  1    1B02_GORGO      P30380 gorilla gor
 38  29  53.7  362  1    1B03_GORGO      P30381 gorilla gor
 39  29  53.7  362  1    1B15_HUMAN      P10317 homo sapien
 40  29  53.7  362  1    1B47_HUMAN      P30487 homo sapien
 41  29  53.7  362  1    1B49_HUMAN      P18464 homo sapien
 42  29  53.7  362  1    1B52_HUMAN      P30489 homo sapien
 43  29  53.7  362  1    1B53_HUMAN      P30490 homo sapien
 44  29  53.7  362  1    1B54_HUMAN      P30491 homo sapien
 45  29  53.7  362  1    1B60_HUMAN      P18465 homo sapien
```

```
                              ALIGNMENTS

RESULT 1
RL17_CHLTR
ID   RL17_CHLTR      STANDARD;      PRT;    122 AA.
AC   P47760;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L17.
GN   RPLQ.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=434/BU / SEROVAR L2;
RX   MEDLINE; 95247702.
RA   GU L.J., WENMAN W.M., REMACHA M., MEUSER R.U., COFFIN J.M., KAUL R.;
RT   "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and
RT   structural analysis.";
RL   J. Bacteriol. 177:2594-2601(1995).
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L33834; AAA74990.1; -.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
DR   PFAM; PF01196; Ribosomal_L17; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   122 AA;    13969 MW;  B8C43F7D CRC32;

Query Match                       63.0%;    Score 34;  DB 1;  Length 122;
Best Local Similarity   54.5%;    Pred. No. 1.3;
Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 RLAIXRIALRY 12
          |||:|:.:||
Db     45 RLAVGRLMVRY 55

RESULT 2
HRMA_PSESY
ID   HRMA_PSESY      STANDARD;      PRT;    375 AA.
AC   Q08370;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HRMA PROTEIN.
GN   HRMA.
OS   Pseudomonas syringae (pv. syringae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U32749; AAC22321.1; -.
DR   TIGR; HI0664; -.
DR   PROSTE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PFAM; PF00664; ABC_membrane; 1.
KW   Hypothetical protein; ATP-binding; Transport; Transmembrane.
FT   TRANSMEM   22   42   POTENTIAL.
FT   TRANSMEM   52   72   POTENTIAL.
FT   TRANSMEM   139   159   POTENTIAL.
FT   TRANSMEM   162   182   POTENTIAL.
FT   TRANSMEM   253   273   POTENTIAL.
FT   TRANSMEM   278   298   POTENTIAL.
FT   NP_BIND   372   378   ATP (POTENTIAL).
SQ   SEQUENCE   552 AA; 61293 MW; C2D5CD9D CRC33;

Query Match   59.3%; Score 32; DB 1; Length 552;
Best Local Similarity   70.0%; Pred. No. 18;
Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   3 LAIXRIALRY 12
          ||:||||||
Db   70 LAVARGALRY 79

RESULT 4
PYRI_SQUAC
ID   PYRI_SQUAC   STANDARD;   PRT;   2242 AA.
AC   Q91437;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CAD PROTEIN (INCLUDES: GLUTAMINE-DEPENDENT CARBAMOYL-PHOSPHATE
DE   SYNTHASE (EC 6.3.5.5); ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2);
DE   DIHYDROOROTASE (EC 3.5.2.3)].
GN   CAD.
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squaliformes; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS, AND SPLEEN;
RX   MEDLINE; 95294021.
RA   HONG J., SALO W.L., ANDERSON P.M.;
RT   "Nucleotide sequence and tissue-specific expression of the
RT   multifunctional protein carbamoyl-phosphate synthetase-aspartate
RT   transcarbamoylase-dihydroorotase (CAD) mRNA in Squalus acanthias.";
RL   J. Biol. Chem. 270:14130-14139(1995).
CC   -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING FOUR
CC   ENZYMATIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC   ATCASE AND DHOASE).
CC   -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +
CC   ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC   -!- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
CC   ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC   -!- CATALYTIC ACTIVITY: DIHYDROOROTATE + H(2)O = N-CARBAMOYL-
CC   L-ASPARTATE.
CC   -!- COFACTOR: DHOASE ACTIVITY REQUIRES A ZINC ATOM.
CC   -!- ENZYME REGULATION: ALLOSTERICALLY REGULATED AND CONTROLLED BY
CC   PHOSPHORYLATION. PRPP IS AN ACTIVATOR WHILE UMP IS AN INHIBITOR
CC   OF THE CPSASE REACTION (BY SIMILARITY).
CC   -!- PATHWAY: FIRST TO THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOHEXAMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- TISSUE SPECIFICITY: PRESENT IN THE TESTIS BUT NOT IN THE LIVER.

RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PSS61;
RA   HUANG H.C., HUTCHESON S.W., COLLMER A.;
RT   "Characterization of the hrp cluster from Pseudomonas syringae pv.
RT   syringae 61 and TnphoA tagging of exported or membrane-spanning Hrp
RT   proteins.";
RL   Mol. Plant Microbe Interact. 4:469-476(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PSS61;
RX   MEDLINE; 94100578.
RA   HEU S., HUTCHESON S.W.;
RT   "Nucleotide sequence and properties of the hrmA locus associated with
RT   the Pseudomonas syringae pv. syringae 61 hrp gene cluster.";
RL   Mol. Plant Microbe Interact. 6:553-564(1993).
CC   -!- FUNCTION: UNKNOWN. MAY SERVE A REGULATORY FUNCTION.
CC   ------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------
DR   EMBL; L14926; AAA16545.1; -.
KW   Hypersensitive response.
SQ   SEQUENCE   375 AA; 41458 MW; 733EEB06 CRC32;

Query Match   61.1%; Score 33; DB 1; Length 375;
Best Local Similarity   58.3%; Pred. No. 7.3;
Matches   7; Conservative   1; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
          ||:||||| ||
Db 174 YRLSITRKTLSY 185

RESULT 3
Y664_HAEIN
ID   Y664_HAEIN   STANDARD;   PRT;   552 AA.
AC   Q57538;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN HI0664.
GN   HI0664.
OS   Haemophilus influenzae.
OC   Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae:
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC   (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC   (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   ------------------------------

CC   -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC   (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC   GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC   -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPASES; THE
CC   DHOASE TO OTHER DHOASES.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC   AMIDOTRANSFERASES.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U18868; AAA74569.1; -.
DR   HSSP; P00479; 3AT1.
DR   PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR   PROSITE; PS00442; GATASE_TYPE_I; 1.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR   PROSITE; PS00483; DIHYDROOROTASE_2; 1.
DR   PROSITE; PS00866; CPSASE_1; 2.
DR   PROSITE; PS00867; CPSASE_2; 2.
DR   PFAM; PF00117; GATase; 1.
DR   PFAM; PF00185; OTCace; 1.
DR   PFAM; PF00289; CPSase_L_chain; 2.
DR   PFAM; PF00744; Dihydroorotase; 1.
DR   PFAM; PF00988; CPSase_sm_chain; 1.
KW   Pyrimidine biosynthesis; Ligase; Transferase; Hydrolase; Zinc;
KW   Allosteric enzyme; Multifunctional enzyme; Phosphorylation.
FT   DOMAIN   1   365   GATASE (GLUTAMINE AMIDOTRANSFERASE).
FT   DOMAIN   366   397   LINKER.
FT   DOMAIN   398   1462   CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE).
FT   DOMAIN   398   937   CPSASE A.
FT   DOMAIN   938   1462   CPSASE B.
FT   DOMAIN   1463   1796   DHOASE (DIHYDROOROTASE).
FT   DOMAIN   1797   1934   LINKER.
FT   DOMAIN   1935   2242   ATCASE (ASPARTATE TRANSCARBAMYLASE).
FT   ACT_SITE   252   252   GATASE (BY SIMILARITY).
FT   ACT_SITE   336   336   GATASE (BY SIMILARITY).
FT   ACT_SITE   338   338   GATASE (BY SIMILARITY).
FT   METAL   1478   1478   ZINC (POTENTIAL).
FT   METAL   1480   1480   ZINC (POTENTIAL).
SQ   SEQUENCE   2242 AA; 249391 MW; 9F38F825 CRC32;

Query Match   59.3%; Score 32; DB 1; Length 2242;
Best Local Similarity   50.0%; Pred. No. 84;
Matches   6; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
             | |:|:|| |||
Db   2109 YLLTLYRVNLRY 2120

RESULT 5
BIOB_MYCTU
ID   BIOB_MYCTU   STANDARD;   PRT;   349 AA.
AC   O06601;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   BIOTIN SYNTHASE (EC 2.8.1.6) (BIOTIN SYNTHETASE).
GN   BIOB OR RV1589 OR MTCY336.15C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PASTEUR;
RA   YU S., JACOBS W.R. JR.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) = BIOTIN.
CC   -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC   FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z95586; CAB09080.1; -.
DR   EMBL; AF041819; AAB96962.1; -.
KW   Biotin biosynthesis; Iron-sulfur; Transferase.
FT   METAL   85   85   IRON-SULFUR (POTENTIAL).
FT   METAL   89   89   IRON-SULFUR (POTENTIAL).
FT   METAL   92   92   IRON-SULFUR (POTENTIAL).
SQ   SEQUENCE   349 AA; 37550 MW; 31B0B9DA CRC32;

Query Match   57.4%; Score 31; DB 1; Length 349;
Best Local Similarity   50.0%; Pred. No. 19;
Matches   6; Conservative   3; Mismatches   3; Indels   3; Gaps   0;

Qy   1 YRLAIXRIALRY 12
             :||||: |: ||
Db   280 FRLALPRTMLRF 291

RESULT 6
SPAS_SALTY
ID   SPAS_SALTY   STANDARD;   PRT;   356 AA.
AC   P40702;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   SURFACE PRESENTATION OF ANTIGENS PROTEIN SPAS.
GN   SPAS.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94008985.
RA   GROISMAN E.A., OCHMAN H.;
RT   "Cognate gene clusters govern invasion of host epithelial cells by
RT   Salmonella typhimurium and Shigella flexneri.";
RL   EMBO J. 12:3779-3787(1993).
CC   -!- FUNCTION: INVOLVED IN A SECRETORY PATHWAY RESPONSIBLE FOR THE
CC   SURFACE PRESENTATION OF DETERMINANTS NEEDED FOR THE ENTRY OF
CC   SALMONELLA SPECIES INTO MAMMALIAN CELLS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC   (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE FLHB/HRPN/YSCU/SPAS FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; X73525; CAA51927.1; -.
DR      PIR; S37356; S37356.
DR      STYGENE; SGI0471; SPAS.
DR      PFAM; PF01312; Bac_export_2; 1.
KW      Virulence; Transmembrane; Inner membrane.
FT      TRANSMEM     29     49       POTENTIAL.
FT      TRANSMEM     72     92       POTENTIAL.
FT      TRANSMEM    132    152       POTENTIAL.
FT      TRANSMEM    179    199       POTENTIAL.
FT      TRANSMEM    261    281       POTENTIAL.
SQ      SEQUENCE    356 AA;  40093 MW;  26881BE1 CRC32;

Query Match             57.4%;  Score 31;  DB 1;  Length 356;
Best Local Similarity   50.0%;  Pred. No. 19;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1  YRLAIXRIALRY 12
           | || : | : ||
Db     70  YSLAVFGIGLKY 81

RESULT   7
YD45_SCHPO      STANDARD;      PRT;    456 AA.
ID  YD45_SCHPO
AC  Q10299;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  HYPOTHETICAL 50.5 KD PROTEIN C22H10.05C IN CHROMOSOME I.
GN  SPAC22H10.05C.
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomyces.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=972;
RA  DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL  Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC  -!- SIMILARITY: TO C.ELEGANS F59A2.4.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; Z69730; CAA93606.1; -.
KW  Hypothetical protein; ATP-binding.
FT  NP_BIND    113    120       ATP (POTENTIAL).
SQ  SEQUENCE   456 AA;  50449 MW;  11934D38 CRC32;

Query Match             57.4%;  Score 31;  DB 1;  Length 456;
Best Local Similarity   50.0%;  Pred. No. 25;
  Matches    5;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1  YRLAIXRIAL 10
           | | : : | : ||
Db    205  YKLSLSRLAL 214

RESULT   8
```

```
PRPR_SALTY      STANDARD;      PRT;    541 AA.
ID  PRPR_SALTY
AC  P74839;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  PROPIONATE CATABOLISM OPERON REGULATORY PROTEIN.
GN  PRPR.
OS  Salmonella typhimurium.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Salmonella.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=LT2;
RX  MEDLINE; 97158691.
RA  HORSWILL A.R., ESCALANTE-SEMERENA J.C.;
RT  "Propionate catabolism in Salmonella typhimurium LT2: two divergently
RT  transcribed units comprise the prp locus at 8.5 centisomes, prpR
RT  encodes a member of the sigma-54 family of activators, and the
RT  prpBCDE genes constitute an operon.";
RL  J. Bacteriol. 179:928-940(1997).
CC  -!- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF THE
CC      PROPIONATE CATABOLISM OPERON.
CC  -!- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR
CC      INTERACTION ATP-BINDING DOMAIN.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; U51879; AAC44813.1; -.
DR  STYGENE; SG?????; PRPR.
DR  PROSITE; PS00675; SIGMA54_INTERACT_1; FALSE_NEG.
DR  PROSITE; PS00676; SIGMA54_INTERACT_2; 1.
DR  PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
DR  PROSITE; PS50045; SIGMA54_INTERACT_4; 1.
DR  PFAM; PF00158; sigma54; 1.
KW  Transcription regulation; DNA-binding; ATP-binding.
FT  DOMAIN     221    464       SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT  NP_BIND    321    330       ATP (POTENTIAL).
FT  DNA_BIND   513    532       H-T-H MOTIF (BY SIMILARITY).
SQ  SEQUENCE   541 AA;  60338 MW;  298D809F CRC32;

Query Match             57.4%;  Score 31;  DB 1;  Length 541;
Best Local Similarity   60.0%;  Pred. No. 30;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1  YRLAIXRIAL 10
           ||| | : ||
Db    389  YRLSILRLTL 398

RESULT   9
YTFM_HAEIN      STANDARD;      PRT;    578 AA.
ID  YTFM_HAEIN
AC  P44038;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  01-NOV-1995 (Rel. 32, Last annotation update)
DE  HYPOTHETICAL PROTEIN HI0698 PRECURSOR.
GN  HI0698.
OS  Haemophilus influenzae.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Haemophilus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RD / KW20;
RX  MEDLINE; 95350630.
```

```
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: STRONG, TO E.COLI YTFM.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U32752; AAC22357.1; -.
DR   TIGR; HI0698; -.
KW   Hypothetical protein; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    578       HYPOTHETICAL PROTEIN HI0698.
SQ   SEQUENCE   578 AA;  65812 MW;  8FCD6AF9 CRC32;

Query Match             57.4%;  Score 31; DB 1;  Length 578;
Best Local Similarity   60.0%;  Pred. No. 32;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YRLAIXRIAL 10
            |  || |:||
Db    147 YKTAISRLAL 156

RESULT 10
MNT_HUMAN
ID   MNT_HUMAN     STANDARD;       PRT;    582 AA.
AC   Q99583;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   MAX BINDING PROTEIN MNT (ROX PROTEIN) (MYC ANTAGONIST MNT).
GN   MNT OR ROX.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE; 97327566.
RA   MERONI G., REYMOND A., ALCALAY M., BORSANI G., TANIGAMI A.,
RA   TONLORENZI R., LO NIGRO C., MESSALI S., ZOLLO M., LEDBETTER D.H.,
RA   BRENT R., BALLABIO A., CARROZZO R.;
RT   "Rox, a novel bHLHzip protein expressed in quiescent cells that
RT   heterodimerizes with Max, binds a non-canonical E box and acts as a
RT   transcriptional repressor.";
RL   EMBO J. 16:2892-2906(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98260677.
RA   NIGRO C.L., VENESIO T., REYMOND A., MERONI G., ALBERICI P.,
RA   CAINARCA S., ENRICO F., STACK M., LEDBETTER D.H., LISCIA D.S.,
RA   BALLABIO A., CARROZZO R.;
RT   "The human ROX Gene: genomic structure and mutation analysis in human
RT   breast tumors.";
RL   Genomics 49:275-282(1998).
CC   -!- FUNCTION: BINDS DNA AS A HETERODIMER WITH MAX AND REPRESSES
CC       TRANSCRIPTION. BINDS TO THE CANONICAL E BOX SEQUENCE
```

```
CC       5'-CACGTG-3' AND, WITH HIGHER AFFINITY, TO 5'-CACGCG-3'
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER WITH MAX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X96401; CAA65265.1; -.
DR   EMBL; Y13440; CAA73851.1; -.
DR   EMBL; Y13441; CAA73851.1; JOINED.
DR   EMBL; Y13442; CAA73851.1; JOINED.
DR   EMBL; Y13443; CAA73851.1; JOINED.
DR   EMBL; Y13444; CAA73851.1; JOINED.
DR   MIM; 603039; -.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
DR   PFAM; PF00010; HLH; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding.
FT   DNA_BIND    222    233       BASIC MOTIF (POTENTIAL).
FT   DNA_BIND    234    270       HELIX-LOOP-HELIX MOTIF (POTENTIAL).
FT   DOMAIN      271    299       LEUCINE-ZIPPER.
SQ   SEQUENCE   582 AA;  62299 MW;  4ED96BBF CRC32;

Query Match             57.4%;  Score 31; DB 1;  Length 582;
Best Local Similarity   60.0%;  Pred. No. 32;
Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 LAIXRIALRY 12
            |:| | ||||
Db    258 LSVLRTALRY 267

RESULT 11
MNT_MOUSE
ID   MNT_MOUSE     STANDARD;       PRT;    591 AA.
AC   O08789; P97349;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   MAX BINDING PROTEIN MNT (ROX PROTEIN) (MYC ANTAGONIST MNT).
GN   MNT OR ROX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 97152466.
RA   HURLIN P.J., QUEVA C., EISENMAN R.N.;
RT   "Mnt, a novel Max-interacting protein is coexpressed with Myc in
RT   proliferating cells and mediates repression at Myc binding sites.";
RL   Genes Dev. 11:44-58(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 97327566.
RA   MERONI G., REYMOND A., ALCALAY M., BORSANI G., TANIGAMI A.,
RA   TONLORENZI R., LO NIGRO C., MESSALI S., ZOLLO M., LEDBETTER D.H.,
RA   BRENT R., BALLABIO A., CARROZZO R.;
RT   "Rox, a novel bHLHzip protein expressed in quiescent cells that
RT   heterodimerizes with Max, binds a non-canonical E box and acts as a
RT   transcriptional repressor.";
RL   EMBO J. 16:2892-2906(1997).
CC   -!- FUNCTION: BINDS DNA AS A HETERODIMER WITH MAX AND REPRESSES
CC       TRANSCRIPTION. BINDS TO THE CANONICAL E BOX SEQUENCE 5'-CACGTG-3'
```

CC   AND, WITH HIGHER AFFINITY, TO 5'-CACGCG-3'
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC   BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER WITH MAX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC   TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC   --------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------
DR   EMBL; U77356; AAB38687.1; -.
DR   EMBL; Y07609; CAA68878.1; -.
DR   MGD; MGI:109150; MNT.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
DR   PFAM; PF00010; HLH; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding.
FT   DNA_BIND   224   235   BASIC MOTIF (POTENTIAL).
FT   DOMAIN   236   272   HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   DOMAIN   273   301   LEUCINE-ZIPPER (POTENTIAL).
FT   CONFLICT   379   379   T -> P (IN REF. 2).
FT   CONFLICT   392   392   V -> A (IN REF. 2).
FT   CONFLICT   402   403   EE -> QQ (IN REF. 2).
FT   CONFLICT   414   414   G -> A (IN REF. 2).
FT   CONFLICT   431   431   V -> A (IN REF. 2).
FT   CONFLICT   465   465   A -> P (IN REF. 2).
FT   CONFLICT   525   525   T -> A (IN REF. 2).
FT   CONFLICT   558   558   G -> A (IN REF. 2).
SQ   SEQUENCE   591 AA; 63311 MW; 16BADBA6 CRC32;

Query Match   57.4%;   Score 31; DB 1; Length 591;
Best Local Similarity   60.0%;   Pred. No. 33;
Matches   6; Conservative   2; Mismatches   2; Indels   2; Gaps   0;

Qy   3 LAIXRIALRY 12
           |:: | ||||
Db   260 LSVLRTALRY 269

RESULT 12
YPD2_CAEEL
ID   YPD2_CAEEL   STANDARD;   PRT;   833 AA.
AC   Q11182;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 93.8 KD PROTEIN C05D11.2 IN CHROMOSOME III
GN   C05D11.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Rhabditinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BRISTOL N2;
RA   DU Z.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: WEAK, TO YEAST VPS16.
CC   --------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------
DR   EMBL; U00048; AAB53823.1; -.
DR   WORMPEP; C05D11.2; CE01132.

KW   Hypothetical protein.
SQ   SEQUENCE   833 AA; 93791 MW; 508777F2 CRC32;

Query Match   57.4%;   Score 31; DB 1; Length 833;
Best Local Similarity   70.0%;   Pred. No. 48;
Matches   7; Conservative   0; Mismatches   0; Indels   3; Gaps   0;

Qy   1 YRLAIXRIAL 10
           | ||| |||
Db   589 YELAISRIPL 598

RESULT 13
CYA6_MOUSE
ID   CYA6_MOUSE   STANDARD;   PRT;   1165 AA.
AC   Q01341;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE   (CA(2+)-INHIBITABLE ADENYLYL CYCLASE).
GN   ADCY6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 9257702.
RA   YOSHIMURA M., COOPER D.M.F.;
RT   "Cloning and expression of a Ca(2+)-inhibitable adenylyl cyclase from
RT   NCB-20 cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:6716-6720(1992).
RN   [2]
RP   SEQUENCE OF 10-1365 FROM N.A.
RX   MEDLINE; 9307607.
RA   PREMONT R.T., JACOBOWITZ O., IYENGAR R.;
RT   "Lowered responsiveness of the catalyst of adenylyl cyclase to
RT   stimulation by GS in heterologous desensitization: a role for
RT   adenosine 3',5'-monophosphate-dependent phosphorylation.";
RL   Endocrinology 131:2774-2784(1992).
CC   -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL
CC   CYCLASE.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC   -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CC   CONCENTRATION RANGE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANT IN HEART BUT WEAKLY DETECTABLE
CC   IN BRAIN, INTESTINE, LUNG, AND SPLEEN.
CC   -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC   -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC   FAMILY.
CC   --------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------
DR   EMBL; M93422; AAA37174.1; -.
DR   EMBL; M96653; AAA37182.1; -.
DR   PIR; A46187; A46187.
DR   HSSP; P19754; 1AWK.
DR   MGD; MGI:87917; ADCY6.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES; 2.
DR   PFAM; PF00211; guanylate_cyc; 2.
KW   Lyase; cAMP synthesis; Transmembrane; Glycoprotein; Duplication.
FT   DOMAIN   1   149   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   150   166   POTENTIAL.
FT   TRANSMEM   179   195   POTENTIAL.
FT   TRANSMEM   212   228   POTENTIAL.

```
FT   TRANSMEM    237    253    POTENTIAL.
FT   TRANSMEM    257    273    POTENTIAL.
FT   TRANSMEM    287    303    POTENTIAL.
FT   DOMAIN      304    670    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    671    688    POTENTIAL.
FT   TRANSMEM    699    715    POTENTIAL.
FT   TRANSMEM    740    756    POTENTIAL.
FT   DOMAIN      757    816    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    817    833    POTENTIAL.
FT   TRANSMEM    836    852    POTENTIAL.
FT   TRANSMEM    894    910    POTENTIAL.
FT   DOMAIN      911   1165    CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    277    277    POTENTIAL.
FT   CARBOHYD    790    790    POTENTIAL.
FT   CARBOHYD    875    875    POTENTIAL.
FT   CONFLICT     47     47    K -> N (IN REF. 2).
FT   CONFLICT     76     76    G -> A (IN REF. 2).
FT   CONFLICT    508    509    GR -> RAG (IN REF. 2).
FT   CONFLICT    737    737    V -> G (IN REF. 2).
FT   CONFLICT    881    881    L -> Q (IN REF. 2).
FT   CONFLICT    990    990    V -> M (IN REF. 2).
FT   SEQUENCE   1165 AA;  130318 MW;  485F7155 CRC32;

Query Match            57.4%;
Best Local Similarity  55.6%;
Matches    5;  Conservative   3;  Mismatches   3;  Indels   1;  Gaps   0;

Qy      4 AIXRIALRY 12
          |:|:||:|
Db    885 AVGRVALKY 893

RESULT 14
CYA6_RAT
ID   CYA6_RAT    STANDARD;    PRT;   1166 AA.
AC   Q03343;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE   (CA(2+)-INHIBITABLE ADENYLYL CYCLASE).
GN   ADCY6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93028552.
RA   PREMONT R.T., CHEN J., MA H.-W., PONNAPALLI M., IYENGAR R.;
RT   "Two members of a widely expressed subfamily of hormone-stimulated
RT   adenylyl cyclases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:9809-9813(1992).
CC   -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL
CC       CYCLASE.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC   -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CC       CONCENTRATION RANGE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC   -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC       FAMILY.
CC   ------------------------------------------------------------------------
```

```
DR   EMBL; L01115; AAA40676.1; -.
DR   EMBL; M96160; AAA40678.1; ALT_INIT.
DR   PIR; A45145; A45145.
DR   HSSP; P19754; 1AWK.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES; 2.
DR   PFAM; PF00211; guanylate_cyc; 2.
KW   Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
FT   DOMAIN        1    149    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    150    166    POTENTIAL.
FT   TRANSMEM    179    195    POTENTIAL.
FT   TRANSMEM    212    228    POTENTIAL.
FT   TRANSMEM    237    253    POTENTIAL.
FT   TRANSMEM    257    273    POTENTIAL.
FT   TRANSMEM    287    303    POTENTIAL.
FT   DOMAIN      304    671    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    672    689    POTENTIAL.
FT   TRANSMEM    700    716    POTENTIAL.
FT   TRANSMEM    741    757    POTENTIAL.
FT   DOMAIN      758    817    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    818    834    POTENTIAL.
FT   TRANSMEM    837    853    POTENTIAL.
FT   TRANSMEM    895    911    POTENTIAL.
FT   DOMAIN      912   1166    CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    791    791    POTENTIAL.
FT   CARBOHYD    876    876    POTENTIAL.
FT   CONFLICT     80     80    K -> E (IN REF. 2).
FT   CONFLICT    130    130    R -> P (IN REF. 2).
FT   CONFLICT    538    538    G -> A (IN REF. 2).
FT   CONFLICT    790    790    I -> L (IN REF. 2).
FT   SEQUENCE   1166 AA;  130506 MW;  AA449BC2 CRC32;

Query Match            57.4%;     Score 31;   DB 1;   Length 1166;
Best Local Similarity  55.6%;     Pred. No. 69;
Matches    5;  Conservative   3;  Mismatches   3;  Indels   1;  Gaps   0;

Qy      4 AIXRIALRY 12
          |:|:||:|
Db    886 AVGRVALKY 894

RESULT 15
LMA3_MOUSE
ID   LMA3_MOUSE    STANDARD;    PRT;   2569 AA.
AC   Q61789; Q61788; Q61966;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   LAMININ ALPHA-3 CHAIN PRECURSOR (FRAGMENT).
GN   LAMA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=LUNG;
RX   MEDLINE; 95949948.
RA   GALLIANO M.-F., ABERDAM D., AGUZZI A., ORTONNE J.-P., MENEGUZZI G.;
RT   "cloning and complete primary structure of the mouse laminin alpha 3
RT   chain. Distinct expression pattern of the laminin alpha 3A and alpha
RT   3B chain isoforms.";
RL   J. Biol. Chem. 270:21820-21826(1995).
RN   [2]
RP   REVISIONS.
RA   ABERDAM D.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [3]
RP   SEQUENCE OF 1052-1770 FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 94281750.
RA   ABERDAM D., GALLIANO M.-F., MATTEI M.-G., PISANI-SPADAFORA A.,
RA   ORTONNE J.-P., MENEGUZZI G.;
RT   "Assignment of mouse nicein genes to chromosomes 1 and 18.";
RL   Mamm. Genome 5:229-233(1994).
RN   [4]
RP   SEQUENCE OF 1052-1770 FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 94363405.
RA   ABERDAM D., AGUZZI A., BAUDOIN C., GALLIANO M.-F., ORTONNE J.-P.,
RA   MENEGUZZI G.;
RT   "Developmental expression of nicein adhesion protein (laminin-5)
RT   subunits suggests multiple morphogenic roles.";
RL   Cell Adhes. Commun. 2:115-129(1994).
CC   -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC       IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC       CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC       WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC   -!- FUNCTION: LAMININ-5 IS THOUGHT TO BE INVOLVED IN (1) CELL ADHESION
CC       VIA INTEGRIN ALPHA-3/BETA-1 IN FOCAL ADHESION AND INTEGRIN ALPHA-
CC       6/BETA-4 IN HEMIDESMOSOMES, (2) SIGNAL TRANSDUCTION VIA TYROSINE
CC       PHOSPHORYLATION OF PP125-FAK AND P80, (3) DIFFERENTIATION OF
CC       KERATINOCYTES (BY SIMILARITY).
CC   -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC       DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC       TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC       COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC       THE ALPHA-3 CHAIN IS A SUBUNIT OF LAMININ-5 (EPILIGRIN/KALININ/
CC       NICEIN), AND POSSIBLY ALSO A COMPONENT OF LAMININ-6 (K-LAMININ)
CC       AND LAMININ-7 (KS-LAMININ).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC       MEMBRANES (MAJOR COMPONENT).
CC   -!- ALTERNATIVE PRODUCTS: THE TWO ISOFORMS A AND B, WHICH DIFFER
CC       IN THEIR N-TERMINAL ARE DERIVED BY ALTERNATIVE SPLICING OF THE
CC       SAME GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE LARGER ISOFORM
CC       B.
CC   -!- TISSUE SPECIFICITY: BASAL MEMBRANE OF THE UPPER ALIMENTARY TRACT
CC       AND URINARY AND NASAL EPITHELIA, SALIVARY GLANDS AND TEETH (BOTH
CC       VARIANTS). ISOFORM A IS PREDOMINANTLY EXPRESSED IN SKIN, HAIR
CC       FOLLICLES AND DEVELOPING NEURONS OF THE TRIGEMINAL GANGLION.
CC       ISOFORM B WAS FOUND IN BRONCHI, ALVEOLI, STOMACH, INTESTINAL
CC       CRYPTS, WHISKER PADS, CNS, TELENCEPHALIC NEUROECTODERM, THALAMUS,
CC       RATHKE'S POUCH, AND PERIVENTRICULAR SUBEPENDYMAL GERMINAL LAYER.
CC   -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC       WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC   -!- DOMAIN: DOMAINS IV AND G ARE GLOBULAR.
CC   -!- SIMILARITY: CONTAINS 6.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV).
CC   -!- SIMILAR TO LAMININ DOMAIN IV).
CC   -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
```

```
DR   EMBL; X84014; CAA58837.1; -.
DR   EMBL; X84013; CAA58836.1; -.
DR   EMBL; L20478; AAA68091.1; -.
DR   HSSP; P02468; 1TLE.
DR   MGD; MGI:99909; LAMA3
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 4.
DR   PFAM; PF00052; laminin_B; 1.
DR   PFAM; PF00053; laminin_EGF; 4.
DR   PFAM; PF00054; laminin_G; 2.

KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER        1      1
FT   SIGNAL        <1     28
FT   CHAIN         29   2569       LAMININ ALPHA-3 CHAIN.
FT   DOMAIN        29    498       DOMAIN IV''.
FT   DOMAIN       499    700       DOMAIN III B.
FT   DOMAIN       546    700       3.5 X LAMININ EGF-LIKE REPEATS.
FT   DOMAIN       546    589       LAMININ EGF-LIKE 1.
FT   DOMAIN       590    639       LAMININ EGF-LIKE 2.
FT   DOMAIN       640    690       LAMININ EGF-LIKE 3.
FT   DOMAIN       691    700       LAMININ EGF-LIKE 4 (N-TERMINAL).
FT   DOMAIN       701    889       LAMININ DOMAIN IV (DOMAIN IV A).
FT   DOMAIN       890   1057       3 X LAMININ EGF-LIKE REPEATS (DOMAIN III
FT                                 A).
FT   DOMAIN       890    922       LAMININ EGF-LIKE 4 (C-TERMINAL).
FT   DOMAIN       923    969       LAMININ EGF-LIKE 5.
FT   DOMAIN       970   1022       LAMININ EGF-LIKE 6.
FT   DOMAIN      1023   1057       LAMININ EGF-LIKE 7 (INCOMPLETE).
FT   DOMAIN      1058   1648       DOMAIN II AND I (HEPTAT REPEATS).
FT   DOMAIN      1649   2569       5 X LAMININ G-LIKE REPEATS (DOMAIN G).
FT   DOMAIN      1649   1825       LAMININ G-LIKE 1.
FT   DOMAIN      1826   1994       LAMININ G-LIKE 2.
FT   DOMAIN      1995   2209       LAMININ G-LIKE 3.
FT   DOMAIN      2210   2385       LAMININ G-LIKE 4.
FT   DOMAIN      2386   2569       LAMININ G-LIKE 5.
FT   DOMAIN      1090   1219       COILED COIL (POTENTIAL).
FT   DOMAIN      1251   1296       COILED COIL (POTENTIAL).
FT   DOMAIN      1327   1404       COILED COIL (POTENTIAL).
FT   DOMAIN      1450   1477       COILED COIL (POTENTIAL).
FT   DOMAIN      1557   1622       COILED COIL (POTENTIAL).
FT   SITE        1513   1515       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     546    553       BY SIMILARITY.
FT   DISULFID     548    560       BY SIMILARITY.
FT   DISULFID     562    571       BY SIMILARITY.
FT   DISULFID     574    587       BY SIMILARITY.
FT   DISULFID     590    605       BY SIMILARITY.
FT   DISULFID     592    612       BY SIMILARITY.
FT   DISULFID     614    623       BY SIMILARITY.
FT   DISULFID     626    637       BY SIMILARITY.
FT   DISULFID     640    652       BY SIMILARITY.
FT   DISULFID     642    659       BY SIMILARITY.
FT   DISULFID     661    670       BY SIMILARITY.
FT   DISULFID     673    688       BY SIMILARITY.
FT   DISULFID     923    932       BY SIMILARITY.
FT   DISULFID     925    939       BY SIMILARITY.
FT   DISULFID     942    951       BY SIMILARITY.
FT   DISULFID     954    967       BY SIMILARITY.
FT   DISULFID     970    982       BY SIMILARITY.
FT   DISULFID     972    991       BY SIMILARITY.
FT   DISULFID     993   1002       BY SIMILARITY.
FT   DISULFID    1005   1020       BY SIMILARITY.
FT   DISULFID    1058   1061       INTERCHAIN (PROBABLE).
FT   DISULFID    1061   1061       INTERCHAIN (PROBABLE).
FT   CARBOHYD     591    591       POTENTIAL.
FT   CARBOHYD     912    912       POTENTIAL.
FT   CARBOHYD    1398   1398       POTENTIAL.
FT   CARBOHYD    1500   1500       POTENTIAL.
FT   CARBOHYD    1571   1571       POTENTIAL.
FT   CARBOHYD    1600   1600       POTENTIAL.
FT   CARBOHYD    1737   1737       POTENTIAL.
FT   CARBOHYD    1819   1819       POTENTIAL.
FT   CARBOHYD    1986   1986       POTENTIAL.
FT   CARBOHYD    2333   2333       POTENTIAL.
FT   CARBOHYD    2509   2509       POTENTIAL.
FT   VARSPLIC       1    842       MISSING (IN ISOFORM A).
FT   VARSPLIC     843    901       MTVLPRLERLCHIRGLHFTETQRLITGEVGLEEASDTGSGPR
FT                                 AHLVEMCACPPDYTGDSC -> MLPAVRWSAWSTGWLWIFG
FT                                 AALGQCLGVGSEQQRVAFLQRPSQNHLQASYMELRPS (IN
FT                                 ISOFORM A).
SQ   SEQUENCE   2569 AA;  282159 MW;  1AF0167A CRC32;
```

Query Match   57.4%; Score 31; DB 1; Length 2569;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches   6; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   3 LAIXRIALRY 12
        |:: || |||
Db   15 LSLFRIVLRY 24

Search completed: February 8, 2000, 00:59:58
Job time: 3787 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 13:17:46 ; Search time 209.03 Seconds
                                             (without alignments)
                                             3.980 Million cell updates/sec

Title: US-08-653-294-19
Perfect score: 54
Sequence: 1 YRLAIXRIALRY 12

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 6934122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   SPTREMBL_12:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
               10: sp_plant:*
               11: sp_rodent:*
               12: sp_virus:*
               13: sp_vertebrate:*
               14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	38	70.4	325	2	Q9XB05	Q9xb05 myxococcus
2	35	64.8	795	2	Q22063	Q22063 caenorhabdi
3	35	64.8	1899	10	Q9XEG1	Q9xeg1 gossypium h
4	34	63.0	141	2	O84514	O84514 chlamydia t
5	34	63.0	309	2	O86347	O86347 mycobacteri
6	33	61.1	142	2	Q9Z7S9	Q9z7s9 chlamydia p
7	32	59.3	264	2	O84015	O84015 chlamydia t
8	32	59.3	310	2	O86606	O86606 streptomyce
9	32	59.3	610	2	P73400	P73400 synechocyst
10	31	57.4	42	11	O54741	O54741 mus musculu
11	31	57.4	190	2	O67313	O67313 aquifex aeo
12	31	57.4	243	2	O85853	O85853 sphingomona
13	31	57.4	407	10	O23555	O23555 arabidopsis
14	31	57.4	469	1	O28589	O28589 archaeoglob
15	31	57.4	910	13	Q9YGE5	Q9yge5 oncorhynchu
16	31	57.4	983	3	O94545	O94545 schizosacch
17	31	57.4	1847	5	P91495	P91495 caenorhabdi
18	30	55.6	171	2	O69220	O69220 azotobacter
19	30	55.6	206	2	Q923X8	Q923x8 pseudomonas
20	30	55.6	211	2	P76343	P76343 escherichia

21	30	55.6	236	2	O88071	streptomyce
22	30	55.6	286	1	O29923	archaeoglob
23	30	55.6	313	2	P97213	clostridium
24	30	55.6	376	1	P95964	sulfolobus
25	30	55.6	377	2	Q9WZD6	thermotoga
26	30	55.6	394	2	Q47265	escherichia
27	30	55.6	430	5	O44725	caenorhabd1
28	30	55.6	441	5	Q26954	trypanosoma
29	30	55.6	497	2	P76767	escherichia
30	30	55.6	573	4	Q9Y4B9	homo sapien
31	30	55.6	880	2	Q9ZD55	rickettsia
32	30	55.6	984	4	O43718	homo sapien
33	30	55.6	1084	2	O83423	treponema p
34	30	55.6	1147	5	Q9XZR7	drosophila
35	30	55.6	1604	10	Q9XI40	arabidopsis
36	29	53.7	35	5	O61237	onchocerca
37	29	53.7	89	5	O19569	homo sapien
38	29	53.7	90	7	O46697	gorilla gor
39	29	53.7	104	8	O11430	avian adeno
40	29	53.7	127	8	Q31747	chara coral
41	29	53.7	133	7	O19189	homo sapien
42	29	53.7	133	12	Q85384	variola vir
43	29	53.7	138	7	O26976	methanobact
44	29	53.7	138	7	O78209	homo sapien
45	29	53.7	144	1	Q9Y9C4	aeropyrum p

ALIGNMENTS

RESULT 1
Q9XB05
ID   Q9XB05   PRELIMINARY;   PRT;   325 AA.
AC   Q9XB05;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MEMBRANE ASSOCIATED PROTEIN.
GN   TAD.
OS   Myxococcus xanthus.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ER-15;
RA   PAITAN Y., ORR E., RON E.Z., ROSENBERG E.;
RT   "Genetic and functional analysis of genes required for the post-
RT   modification of the polyketide antibiotic TA of Myxococcus xanthus.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ132503; CAB46503.1; -.
SQ   SEQUENCE   325 AA; 35985 MW; 4CC64E85 CRC32;

Query Match   70.4%; Score 38; DB 2; Length 325;
Best Local Similarity   58.3%;   Pred. No. 2.9;
Matches   7; Conservative   1; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
            ||| : ||
Db   202 YRLTVDRFPLRY 213

RESULT 2
Q22063
ID   Q22063   PRELIMINARY;   PRT;   795 AA.
AC   Q22063; Q93489;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T01C3.10 PROTEIN.
GN   T01C3.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WILD A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; 281061; CAB02939.1; -.
DR   EMBL; 278413; CAB02939.1; JOINED.
DR   EMBL; 278413; CAB01667.1; -.
DR   EMBL; 281061; CAB01667.1; JOINED.
DR   HSSP; P19491; 1GR2.
DR   PFAM; PF00060; lig_chan; 1.
SQ   SEQUENCE   795 AA;   89703 MW; DD722166 CRC32;

Query Match   64.8%; Score 35; DB 5; Length 795;
Best Local Similarity   50.0%; Pred. No. 31;
  Matches   6; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
        || :: |:| ||
Db   6 YRTSLRRLATRY 17

RESULT 3
Q9XEG1
ID   Q9XEG1
AC   Q9XEG1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE CALLOSE SYNTHASE CATALYTIC SUBUNIT.
GN   CFL1.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. TEXAS MARKER-1; TISSUE=PRIMARY-STAGE COTTON FIBER;
RA   CUI X., SHIN H., BROWN R.M.;
RT   "Cotton CFL1 gene shows homology to the yeast beta-1,3-glucan synthase
RT   subunit FKS1.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF085717; AAD25952.1; -.
SQ   SEQUENCE   1899 AA;   218627 MW; E695145F CRC32;

Query Match   64.8%; Score 35; DB 10; Length 1899;
Best Local Similarity   63.6%; Pred. No. 79;
  Matches   7; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIXRIALRY 12
        || |: || ||
Db 554 RLAVSRIFLRF 564

RESULT 4
O84514
ID   O84514
AC   O84514;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L17.
GN   RL17.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;

RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL; AE001323; AAC68107.1; -.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
DR   PFAM; PF01196; Ribosomal_L17; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   141 AA;   16152 MW; 2570FFF7 CRC32;

Query Match   63.0%; Score 34; DB 2; Length 141;
Best Local Similarity   54.5%; Pred. No. 8.3;
  Matches   6; Conservative   3; Mismatches   3; Indels   2; Gaps   0;

Qy   2 RLAIXRIALRY 12
        ||| |: :||
Db   64 RLAVGRLMVRY 74

RESULT 5
O86347
ID   O86347
AC   O86347;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)   PRELIMINARY;   PRT;   309 AA.
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 33.5 KD PROTEIN.
GN   RV2776C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY III C.E., TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   PARKHILL J.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL008967; CAA15591.1; -.
DR   HSSP; P33164; 2PIA.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PFAM; PF00111; fer2; 1.
DR   PFAM; PF00175; oxidored_fad; 1.
KW   Hypothetical protein; Iron-sulfur.
SQ   SEQUENCE   309 AA;   33517 MW; B152B590 CRC32;

Query Match   63.0%; Score 34; DB 2; Length 309;
Best Local Similarity   77.8%; Pred. No. 19;
  Matches   7; Conservative   1; Mismatches   1; Indels   1; Gaps   0;

Qy   1 YRLAIXRIA 9

```
           |||:||:|||
Db   65 YRIAIRRIA 73

RESULT 6
ID   Q9Z7S9      PRELIMINARY;      PRT;    142 AA.
AC   Q9Z7S9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L17.
GN   RL17.
OS   Chlamydia pneumoniae.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RA   KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA   GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT   "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL; AE001647; AAD18764.1; -.
DR   PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   142 AA;  16400 MW;  4839EC84 CRC32;

Query Match                    61.1%;  Score 33;  DB 2;  Length 142;
Best Local Similarity          54.5%;  Pred. No. 14;
Matches    6;  Conservative    3;  Mismatches    3;  Indels    2;  Gaps    0;

Qy    2 RLAIXRIALRY 12
            |:|| |: :||
Db   64 RIAIGRLMVRY 74

RESULT 7
ID   O84015      PRELIMINARY;      PRT;    264 AA.
AC   O84015;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 30.0 KD PROTEIN.
GN   YBBP.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001276; AAC67602.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   264 AA;  29971 MW;  C0AD12DB CRC32;

Query Match                    59.3%;  Score 32;  DB 2;  Length 264;
Best Local Similarity          70.0%;  Pred. No. 42;
Matches    7;  Conservative    1;  Mismatches    1;  Indels    2;  Gaps    0;


Qy    2 RLAIXRIALR 11
            |||:|| ||
Db   83 RLALSRIRLR 92

RESULT 8
ID   O86606      PRELIMINARY;      PRT;    310 AA.
AC   O86606;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PUTATIVE TRANSPOSASE.
GN   SC3A7.05C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   OLIVER K., HARRIS D.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE: 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol 21:77-96(1996).
DR   EMBL; AL031155; CAA20068.1; -.
SQ   SEQUENCE   310 AA;  33156 MW;  5AD2CED4 CRC32;

Query Match                    59.3%;  Score 32;  DB 2;  Length 310;
Best Local Similarity          63.6%;  Pred. No. 50;
Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    2 RLAIXRIALRY 12
            |||:| | ||
Db  249 RLAVFRWATRY 259

RESULT 9
ID   P73400      PRELIMINARY;      PRT;    610 AA.
AC   P73400;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ABC TRANSPORTER.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE: 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
```

RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL: D90906; BAA17440.1; -.
HSSP; P13569; 1NBD.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PFAM; PF00005; ABC_tran; 1.
SQ   SEQUENCE   610 AA;   67028 MW; 835D2045 CRC32;

Query Match   59.3%; Score 32; DB 2; Length 610;
Best Local Similarity   70.0%; Pred. No. 1e+02;
Matches   7; Conservative   1; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIXRIALR 11
          |||| | |:|
Db   508 RLAIARAAMR 517

RESULT 10
O54741
ID   O54741   PRELIMINARY;   PRT;   42 AA.
AC   O54741;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   LAMININ, ALPHA 3 (LAMININ 5 ALPHA3C CHAIN) (FRAGMENT).
GN   LAMA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SV129;
RX   MEDLINE; 97400527.
RA   FERRIGNO O., VIROLLE T., GALLIANO M.F., CHAUVIN N., ORTONNE J.P.,
RA   MENEGUZZI G., ABERDAM D.;
RT   "Murine laminin alpha3A and alpha3B isoform chains are generated by
RT   usage of two promoters and alternative splicing.";
RL   J. Biol. Chem. 272:20502-20508(1997).
DR   EMBL; Y08850; CAA70073.1; -.
DR   MGD; MGI:99909; Lama3.
FT   NON_TER   42   42
SQ   SEQUENCE   42 AA;   4872 MW; 42FE6150 CRC32;

Query Match   57.4%; Score 31; DB 11; Length 42;
Best Local Similarity   60.0%; Pred. No. 10;
Matches   6; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   3 LAIXRIALRY 12
          |:: || |||
Db   22 LSLFRIVLRY 31

RESULT 11
O67313
ID   O67313   PRELIMINARY;   PRT;   190 AA.
AC   O67313;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 22.0 KD PROTEIN.
GN   AQ_1277.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,

RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000732; AAC07278.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   190 AA;   22041 MW; EFB61F50 CRC32;

Query Match   57.4%; Score 31; DB 2; Length 190;
Best Local Similarity   54.5%; Pred. No. 49;
Matches   6; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALR 11
          : ||| ::|||
Db   43 HNLAIQKVALR 53

RESULT 12
O85853
ID   O85853   PRELIMINARY;   PRT;   243 AA.
AC   O85853;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 26.5 KD PROTEIN.
OS   Sphingomonas aromaticivorans.
OG   Plasmid pNL1.
OC   Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
OC   Sphingomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F199;
RA   ROMINE M.F., STILLWELL L.C., WONG K.-K., THURSTON S.J., SISK E.C.,
RA   SENSEN C.W., GAASTERLAND J.D., SAFFER J.D., FREDRICKSON J.K.;
RT   "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas
RT   aromaticivorans strain F199.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF079317; AAD03868.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   243 AA;   26455 MW; 40CDFBF4 CRC32;

Query Match   57.4%; Score 31; DB 2; Length 243;
Best Local Similarity   54.5%; Pred. No. 63;
Matches   6; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   2 RLAIXRIALRY 12
          ||| |: :||
Db   5 RLAQRRVTIRY 15

RESULT 13
O23555
ID   O23555   PRELIMINARY;   PRT;   407 AA.
AC   O23555;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE SERINE PROTEASE-LIKE PROTEIN.
GN   DL4585C.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,
RA   KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA   PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA   JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA   SCHUELLER C., CHALWATZIS N.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97342; CAB46052.1; -.
KW   Protease.
SQ   SEQUENCE   407 AA; 46703 MW; 9B2D1556 CRC32;

Query Match   57.4%; Score 31; DB 10; Length 407;
Best Local Similarity   58.3%; Pred. No. 1.1e+02;
Matches   7; Conservative   1; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
              ||| |||| :
Db   368 YRLGNNRIALSF 379

RESULT 14
028589
ID   028589;   PRELIMINARY;   PRT;   469 AA.
AC   028589;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   AF1684.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE000987; AAB89555.1; -.
DR   TIGR; AF1684; -.
KW   Hypothetical protein.
SQ   SEQUENCE   469 AA; 53468 MW; 9D430846 CRC32;

Query Match   57.4%; Score 31; DB 1; Length 469;
Best Local Similarity   58.3%; Pred. No. 1.3e+02;
Matches   7; Conservative   1; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
              || : |||| |
Db   12 YRKGMKRIALVY 23

RESULT 15

Q9YGE5
ID   Q9YGE5   PRELIMINARY;   PRT;   910 AA.
AC   Q9YGE5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RETINOBLASTOMA 1.
GN   RB1.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Protacanthopterygii;
OC   Salmoniformes; Salmonidae; Oncorhynchus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99132290.
RA   BRUNELLI J.P., THORGAARD G.H.;
RT   "Sequence, expression and genetic mapping of a rainbow trout
RT   retinoblastoma cDNA.";
RL   Gene 226:175-180(1999).
DR   EMBL; AF102861; AAD13390.1; -.
DR   HSSP; P06400; 1GUX.
SQ   SEQUENCE   910 AA; 103531 MW; CE74234B CRC32;

Query Match   57.4%; Score 31; DB 13; Length 910;
Best Local Similarity   50.0%; Pred. No. 2.5e+02;
Matches   6; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIXRIALRY 12
              |||| |: : |
Db   644 YRLAYMRLKMLY 655

Search completed: February 8, 2000, 13:17:48
Job time: 32497 sec

THIS PAGE BLANK (USPTO)

```
OM of: US-08-653-294-19 to: GenEmbl:*      out_format : pfs

Date: Feb 8, 2000  4:43 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1/1/USPTO_spool/US08653294/runat_04022000_160701_15779/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-19
Query length: 12
Database: GenEmbl:*
Database sequences: 821193
Database length: 1518192014
Search time (sec): 11370.480000
```

```
score_list:
Sequence              Strd  Orig   ZScore  EScore    Len     | Documentation
gb_in1:DMRNA3          +    44.00  158.56  0.8265    1050     | V00238   Three Drosophila melanc
gb_htg7:AC017966       +    44.00  139.81  9.17      8849     | AC017966 Drosophila melanogast
gb_htg4:AC010564       +    44.00  115.26  213.47    143914   ! AC010564 Drosophila melanoga
gb_in2:AC005847        -    44.00  109.78  431.34    268369   ! AC005847 Drosophila melanoga
gb_ba1:MXA132503       +    38.00  120.94  103.03    4513     ! AJ132503 Myxococcus xanthus ta
gb_pr4:U72788          +    38.00  105.74  723.80    25383    ! U72788   Homo sapiens cosmid cl
gb_ba1:U00020          +    38.00  102.44  1.1e+03   36947    ! U00020   Mycobacterium leprae c
gb_ba1:MLCB2533        +    38.00  101.68  1.2e+03   40245    ! AL035310 Mycobacterium leprae
gb_ba1:U00017          -    38.00  101.27  1.3e+03   42157    ! U00017   Mycobacterium leprae c
gb_pl2:ATAC005560      +    38.00   93.40  3.5e+03   103125   ! AC005560 Arabidopsis thalian
gb_in1:CNS01DFM        -    38.00   87.96  7.1e+03   191442   ! AL132800 Homo sapiens chromo
gb_in1:DVU57315        -    37.00  124.37  66.35     1911     ! U57315   Drosophila virilis LAMA
gb_pr3:HS782L23        +    37.00   87.65  7.4e+03   123925   ! AL035416 Human DNA sequence
gb_htg4:AC011241       -    37.00   84.01  1.2e-04   187493   ! AC011241 Homo sapiens chromo
gb_ba2:AF026541        +    36.00  121.31  98.25     1692     ! AF026541 Mycobacterium tubercu
gb_pat:E14720          +    36.00  121.13  100.61    1728     ! E14720   Procine mRNA for alpha-
gb_om:D86723           -    36.00  120.59  107.73    1836     ! D86723   Rattus norvegicus prost
gb_ro:RATPES2A         +    36.00  117.10  168.59    2730     ! L11611   Rattus norvegicus prost
gb_htg5:AC013927       -    36.00  103.52  962.01    12771    ! AC013927 Drosophila melanogas
gb_in1:CEW02D9         +    36.00   94.64  3.0e+03   35024    ! Z81137   Caenorhabditis elegans
gb_htg4:AC012150       -    36.00   94.05  3.0e+03   37459    ! AC012150 Homo sapiens chromos
gb_ba1:MTCY05A6        -    36.00   93.78  3.4e+03   38631    ! Z96072   Mycobacterium tubercul
gb_ba1:MSGB13GS        +    36.00   92.85  3.8e+03   42923    ! L78823   Mycobacterium leprae c
gb_ba1:MLCL458         +    36.00   92.67  3.9e+03   43839    ! AL049478 Mycobacterium leprae
gb_htg3:AC009588       -    36.00   92.12  4.2e+03   46632    ! AC009588 Homo sapiens chromos
gb_htg2:AC004378       +    36.00   86.60  8.4e+03   87319    ! AC004378 Drosophila melanogas
gb_htg6:AC011787       +    36.00   86.52  8.5e+03   88178    ! AC011787 Homo sapiens clone F
gb_htg2:AC005978       -    36.00   86.09  9.0e+03   92586    ! AC005978 Drosophila melanogas
gb_pr3:HS614C15        +    36.00   84.95  1.0e+04   105325   ! AL035250 Human DNA sequence
gb_htg5:AC009909       -    36.00   84.82  1.1e+04   106956   ! AC009909 Drosophila melanoga
gb_htg6:AC007807       +    36.00   83.06  1.3e+04   130536   ! AC007807 Drosophila melanoga
gb_htg7:AC013772       +    36.00   82.27  1.5e+04   142796   ! AC013772 Homo sapiens clone
gb_htg2:AC007924       -    36.00   82.17  1.5e+04   144526   ! AC007924 Homo sapiens clone
gb_htg6:AC012364       +    36.00   81.85  1.6e+04   149898   ! AC012364 Homo sapiens clone
gb_htg3:AC008521       +    36.00   81.59  1.6e+04   154276   ! AC008521 Homo sapiens chromo
gb_htg5:AC011145       -    36.00   81.48  1.6e+04   156243   ! AC011145 Homo sapiens clone
gb_vi:AF157706         +    36.00   81.16  1.7e+04   162114   ! AF157706 Human herpesvirus 6
gb_htg4:AC008922       +    36.00   80.78  1.8e+04   169299   ! AC008922 Homo sapiens chromo
gb_htg3:AC009657       -    36.00   80.42  1.9e+04   176365   ! AC009657 Homo sapiens chromo
gb_htg4:AC010981       +    36.00   80.10  1.9e+04   182860   ! AC010981 Homo sapiens chromo
gb_htg1:HSJ492J12      +    36.00   79.97  2.0e+04   185485   ! AL121919 Homo sapiens chromo
gb_pr4:AC004168        -    36.00   79.97  2.0e+04   185566   ! AC004168 Homo sapiens chromo
gb_pr2:CNS01DS2        +    36.00   79.56  2.1e+04   194430   ! AL121652 BAC sequence fro
gb_htg4:AC008810       -    36.00   78.70  2.3e+04   214385   ! AC008810 Homo sapiens chr
gb_pr2:CNS00008        +    36.00   78.51  2.4e+04   218956   ! AL049831 Human chromosome
```

```
seq_name: gb_in1:DMRNA3

seq_documentation_block:
LOCUS       DMRNA3          1050 bp    DNA            INV       31-MAR-1992
DEFINITION  Three Drosophila melanogaster genes for transfer RNAs (Glu
            specific).
ACCESSION   V00238 J01146
VERSION     V00238.1  GI:8458
KEYWORDS    transfer RNA; transfer RNA-Glu.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1050)
AUTHORS     Hosbach,H.A., Silberklang,M. and McCarthy,B.J.
TITLE       Evolution of a D. melanogaster glutamate tRNA gene cluster
JOURNAL     Cell 21 (1), 169-178 (1980)
MEDLINE     81001852
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            1..1050
                     /gene="tRNA:E4:62Aa"
                     /note="transfer RNA:glu4:62Aa"
                     /allele=""
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            25..96
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            262..333
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
     tRNA            464..535
                     /gene="tRNA:E4:62Aa"
                     /note="tRNA Glu"
                     /db_xref="FlyBase:FBgn0011851"
BASE COUNT      324 a    190 c    213 g    323 t
ORIGIN

alignment_scores:
                    Quality:   44.00            Length:      12
                     Ratio:     4.000             Gaps:       0
         Percent Similarity:   91.667    Percent Identity:   66.667

alignment_block:
US-08-653-294-19 x DMRNA3

Align seg 1/1   to: DMRNA3   from: 1  to: 1050

       1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr   12
         ||||||:::::::::   ||||||||||||||||||
     922 TACAGAATGTCGATGCAACGCATTGCTTACGTTAT  957

seq_name: gb_htg7:AC017966

seq_documentation_block:
LOCUS       AC017966        8849 bp    DNA      HTG      09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017966
VERSION     AC017966.1  GI:6553224
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
```

REFERENCE   1 (bases 1 to 8849)
AUTHORS   Adams,M. and Venter,J.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA

COMMENT   This sequence was identified as CDM:10212817 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.

FEATURES   Location/Qualifiers
     source   1..8849
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"

BASE COUNT   2604 a   1735 c   1884 g   2626 t
ORIGIN

alignment_scores:
         Quality: 44.00   Length: 12
         Ratio: 4.000   Gaps: 0
Percent Similarity: 91.667   Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x AC017966

Align seg 1/1   to: AC017966   from: 1 to: 8849

     1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
       |||||::::::::::::: ||||||||||||||
   844 TACAGAATGTCGATGCAACGCATTGCTTTACGTTAT 879

seq_name: gb_htg4:AC010564

LOCUS   AC010564   143914 bp   DNA   HTG   16-OCT-1999
DEFINITION Drosophila melanogaster chromosome 3L/62A1 clone RPCI98-2701, ***
            SEQUENCING IN PROGRESS ***, 89 unordered pieces.
ACCESSION   AC010564
VERSION   AC010564.4   GI:5917942
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 143914)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
            Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
            Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
            Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
            Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
            Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P.,
            Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
            Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 143914)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (16-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Sep 22, 1999 this sequence version replaced gi:5916428.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 89 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
*
*   1   831: contig of   831 bp in length
*   832   1645: contig of   814 bp in length
*   1646   2519: contig of   874 bp in length
*   2520   3350: contig of   831 bp in length
*   3351   4206: contig of   856 bp in length
*   4207   5469: contig of   1263 bp in length
*   5470   6313: contig of   844 bp in length
*   6314   7273: contig of   960 bp in length
*   7274   8101: contig of   828 bp in length
*   8102   8941: contig of   840 bp in length
*   8942   10144: contig of   1203 bp in length
*   10145   11019: contig of   875 bp in length
*   11020   11825: contig of   806 bp in length
*   11826   13085: contig of   1260 bp in length
*   13086   13937: contig of   852 bp in length
*   13938   14752: contig of   815 bp in length
*   14753   15972: contig of   1220 bp in length
*   15973   16571: contig of   599 bp in length
*   16572   17184: contig of   613 bp in length
*   17185   18018: contig of   834 bp in length
*   18019   18837: contig of   819 bp in length
*   18838   20391: contig of   1554 bp in length
*   20392   21252: contig of   861 bp in length
*   21253   21970: contig of   718 bp in length
*   21971   23026: contig of   1056 bp in length
*   23027   24541: contig of   1515 bp in length
*   24542   25726: contig of   1185 bp in length
*   25727   26570: contig of   844 bp in length
*   26571   27681: contig of   1111 bp in length
*   27682   29230: contig of   1549 bp in length
*   29231   30635: contig of   1405 bp in length
*   30636   31757: contig of   1122 bp in length
*   31758   32899: contig of   1142 bp in length
*   32900   34092: contig of   1193 bp in length
*   34093   34928: contig of   836 bp in length
*   34929   35705: contig of   777 bp in length
*   35706   36891: contig of   1186 bp in length
*   36892   38403: contig of   1512 bp in length
*   38404   39938: contig of   1535 bp in length
*   39939   41157: contig of   1219 bp in length
*   41158   42005: contig of   848 bp in length
*   42006   43465: contig of   1460 bp in length
*   43466   44890: contig of   1425 bp in length
*   44891   45926: contig of   1036 bp in length
*   45927   47498: contig of   1572 bp in length
*   47499   49100: contig of   1602 bp in length
*   49101   50371: contig of   1271 bp in length
*   50372   51976: contig of   1605 bp in length
*   51977   54050: contig of   2074 bp in length
*   54051   56027: contig of   1977 bp in length
*   56028   57411: contig of   1384 bp in length
*   57412   58815: contig of   1404 bp in length
*   58816   60132: contig of   1317 bp in length
*   60133   61360: contig of   1228 bp in length
*   61361   63494: contig of   2134 bp in length
*   63495   64634: contig of   1140 bp in length
*   64635   65855: contig of   1221 bp in length
*   65856   66834: contig of   979 bp in length
*   66835   68045: contig of   1211 bp in length
*   68046   68873: contig of   828 bp in length

```
*   68874: contig of 1146 bp in length
*   70020: contig of 2304 bp in length
*   72324: contig of 1436 bp in length
*   73760: contig of 2075 bp in length
*   75835: contig of 2555 bp in length
*   78390: contig of 2451 bp in length
*   80841: contig of 1271 bp in length
*   82112: contig of 1941 bp in length
*   84053: contig of 2058 bp in length
*   86111: contig of 1378 bp in length
*   87489: contig of 1620 bp in length
*   89109: contig of 2349 bp in length
*   91458: contig of 1450 bp in length
*   92908: contig of 2430 bp in length
*   95338: contig of 1758 bp in length
*   97096: contig of 2996 bp in length
* 100092: contig of 1981 bp in length
* 102073: contig of 2425 bp in length
* 104498: contig of 3011 bp in length
* 107509: contig of 2090 bp in length
* 109599: contig of 2807 bp in length
* 112406: contig of 2799 bp in length
* 115205: contig of 3113 bp in length
* 118318: contig of 3568 bp in length
* 121886: contig of 3336 bp in length
* 125222: contig of 4117 bp in length
* 129339: contig of 3473 bp in length
* 132812: contig of 5081 bp in length
* 137893: contig of 6022 bp in length.

FEATURES        Location/Qualifiers
     source          1..143914
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3L/62A1"
                     /clone="RPCI98-27O1"
BASE COUNT    38720 a  33523 c  32072 g  39042 t  557 others
ORIGIN

alignment_scores:
     Quality:           44.00           Length:       12
     Ratio:             4.000           Gaps:          0
Percent Similarity:    91.667     Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x AC010564

Align seq 1/1  to: AC010564   from: 1  to: 143914

       1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         |||||:::::::::  |||||||||||||||||||
   58362 TACAGAATGTCGATGCAACGCCATTGCTTTACGTTAT 58397

seq_name: gb_in2:AC005847

seq_documentation_block:
LOCUS       AC005847   268369 bp   DNA        INV    22-OCT-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS03179 (D226), DS06357
            (D230), DS03404 (D231), DS06962 (D232), and DS07291 (D240)),
            complete sequence.
ACCESSION   AC005847 AC004331 AC004336 AC004436 AC004334 AC004488 AC004489
VERSION     AC005847.1  GI:3779012
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clones DS03179
            (D226), DS06357 (D230), DS03404 (D231), DS06962 (D232), and DS07291
            (D240) DNA).
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 268369)
AUTHORS     Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
```

```
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE       Sequencing of Drosophila chromosome 3L, region 61F3-62A2
JOURNAL     Unpublished (1997)
REFERENCE   2 (bases 1 to 268369)
AUTHORS     Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE       Direct Submission
JOURNAL     Submitted (22-OCT-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            Library locations: 11-34, 21-67, 44-36, 73-50, 92-76.
FEATURES        Location/Qualifiers
     source          1..268369
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3L"
                     /map="61F3-62A2"
                     /clone="P1s DS03179 (D226), DS06357 (D230), DS03404
                     (D231), DS06962 (D232), and DS07291 (D240)"
                     /note="These five P1s were completed as a project.  The
                     coordinates are the P1 ends for each P1. DS03404
                     (D231)extends from bp 1 to bp 80,732.  DS03179 (D226)
                     extends from bp 32,734 to bp 117,954.  DS06357 (D230)
                     extends from bp 88,745 to bp 169,482.  DS06962 (D232)
                     extends from bp 127,117 to bp 209,427.  DS07291 (D240)
                     extends from bp 194,759 to bp 268,369."
BASE COUNT    71701 a  61029 c  61736 g  73903 t
ORIGIN

alignment_scores:
     Quality:           44.00           Length:       12
     Ratio:             4.000           Gaps:          0
Percent Similarity:    91.667     Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x AC005847/rev

Align seq 1/1  to reverse of: AC005847   from: 1  to: 268369

       1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         |||||:::::::::  |||||||||||||||||||
   46162 TACAGAATGTCGATGCAACGCCATTGCTTTACGTTAT 46127

seq_name: gb_bal:MXA132503

seq_documentation_block:
LOCUS       MXA132503   4513 bp   DNA           BCT    01-JUL-1999
DEFINITION  Myxococcus xanthus taB, taC, taD, taE and taF genes.
ACCESSION   AJ132503
VERSION     AJ132503.1  GI:5419951
KEYWORDS    ACP synthase; acyl carier protein (ACP); membrane associated
            protein; taB gene; taC gene; taD gene; taE gene; TAF gene.
SOURCE      Myxococcus xanthus
ORGANISM    Myxococcus xanthus
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
```

```
REFERENCE     Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
AUTHORS       1 (bases 1 to 4513)
TITLE         Paitan,Y., Orr,E., Ron,E.Z. and Rosenberg,E.
              Genetic and functional analysis of genes required for the
              post-modification of the polyketide antibiotic TA of Myxococcus
              xanthus
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 4513)
AUTHORS       Paitan,Y.
TITLE         Direct Submission
JOURNAL       Submitted (25-JAN-1999) Paitan Y., Molecular Microbiology and
              Biotechnology, Tel Aviv University, G S Wise Faculty of life
              Sciences, Tel Aviv University, Ramat Aviv, 69978, ISRAEL
FEATURES      Location/Qualifiers
source        1 .4513
              /organism="Myxococcus xanthus"
              /strain="ER-15"
              /db_xref="taxon:34"
gene          431..670
              /gene="taB"
CDS           431..670
              /gene="taB"
              /codon_start=1
              /transl_table=11
              /product="acyl carrier protein (ACP)"
              /protein_id="CAB46501.1"
              /db_xref="GI:5419952"
              /translation="MDKRIIFDIVTSSVREVVPELESHPFEPEDDLVGLGANSLDRAE
              IVNLTLEKALNIPRVELIDAKTIGGLVDVLHARL"
gene          680..1942
              /gene="taC"
CDS           680..1942
              /gene="taC"
              /note="ACP synthase (KAS III) homologue"
              /codon_start=1
              /transl_table=11
              /product="TaC protein"
              /protein_id="CAB46502.1"
              /db_xref="GI:5419953"
              /translation="MGPVGIEAMNAYCGIARLDVLQLATHRGLDTSRFANLLMEEKTV
              PLPEDPVTYGVNAARPIIDGLTAAERDSIELLVACTESSFDFGKAMSTYLHQHLGLS
              RNCRLIELKSACYSGVAGLQMAVNFILSGVSPGAKALVVASDLSRFSIAEGDASTED
              WSFAEPSSGAGAVAMLVSDTPRVFRVDVGANGYYGEVMDTCRPVADSEAGDADLSLL
              SYLDCCENAFREYIRRVPAANYAESFGYLAFHTPFGGMVKGAHRTMMRKFSGKNRGDI
              EADFQRRVAPGLTYCQRVGNIMGATMALSLLGTIDHGDFATAKRIGCFSYGSGCSSEF
              FSGVVTEEGQQRQRALGLGEALGRRQQLSMPDYDALLKGNGLVRFGTRNAELDFGVVG
              SIRPGGWGRPLLFLSAIRDFHRDIQWIS"
gene          1965..2942
              /gene="taD"
CDS           1965..2942
              /gene="taD"
              /codon_start=1
              /transl_table=11
              /product="membrane associated protein"
              /protein_id="CAB46503.1"
              /db_xref="GI:5419954"
              /translation="MSSVATAVPLTARDSAVSRRLRITPSMCGQTSLFAGQIGDWAWD
              TVSRLCGTDVLRATNASGAPTYLAFYIFRIRGTPALHPGALRFGDTLDVTSKAYNFGS
              ESVLTVHRICKTAEGGAPEADAFGHEELYEQPQFGRIYAETFNRWITRSDGKSNESLI
              KSSPVGFQYAHLPLPDEYSPRRAYGDARARGTFHDVDSAEYRLTVDRFPLRYAVDVI
              RDVNGVGLIYFASYFSMVDWAIWQLARHQGRSEQAFLSRVVLDQQLCFLGNAALDTTF
              DIDVQHWERVGGGGELFNVKMREGAQGRDIAVATVKVRFDAASEGGRG"
gene          2947..3198
              /gene="taE"
CDS           2947..3198
              /gene="taE"
              /codon_start=1
              /transl_table=11
              /product="acyl carrier protein (ACP)"
              /protein_id="CAB46504.1"
              /db_xref="GI:5419955"
              /translation="MTDEQIRGVVHQSIVRVLPRVRSNEIAGHLNLRELGADSVDRVE
              ILTSILDSLRLQKTPLAKFADIRNIDALVAFLAGEVAGG"
```

```
gene          3318..4442
              /gene="taF"
CDS           3318..4442
              /gene="taF"
              /note="TaF protein"
              /codon_start=1
              /transl_table=11
              /product="ACP synthase (KAS III) homologue"
              /protein_id="CAB46505.1"
              /db_xref="GI:5419956"
              /translation="MMQERGVALPEDPVTNAVNAARPILDAMSPEARERIELLVTSS
              ESGVDFSKSISSAHEHLGLSRHCRFLEVKQACYAATGALQLALGYIASGVSPGAKAL
              VIATDVTLVDESGIYSEPAMGTGGVAVLLGDEPRVMKMDLGAFGNYSYDVFDTARPSP
              EIDIGDVDRSLFTYLDCLKHSFAAYGRRVDGVDFVSTFDYLAMHTPFAGLVKAGHRKM
              MRELTPCDVDEIEADFGRRVKPSLQYPSLVGNLCSGSVVLSLCSIIDTIKPERSARVG
              MFSTGSGCSSEFFSGVIGPESVSALAGLDIGGHLRGRRQLTFDQYVELLKENLRCLVP
              TKNRDVDVERYLPLVTRTASRPRMALRRVVDYHRQYEWV"
BASE COUNT     734 a   1461 c   1534 g    784 t
ORIGIN
```

alignment_scores:
```
              Quality:       38.00        Length:        12
              Ratio:          3.800        Gaps:           0
Percent Similarity:  83.333    Percent Identity:  58.333
```

alignment_block:
```
US-08-653-294-19 x MXA132503

Align seg 1/1 to: MXA132503  from: 1 to: 4513

         1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
           ||||||||||::::::::  |||:::  ||||||||||
      2568 TACCGGCTGACCGTGGACCGCTTCCCGCTGCGCTAC 2603
```

seq_name: gb_pr4:U72788

seq_documentation_block:
```
LOCUS         U72788        25383 bp    DNA         PRI       27-APR-1999
DEFINITION    Homo sapiens cosmid clone U169D2 from Xp22.1-22.2, complete
              sequence.
ACCESSION     U72788
VERSION       U72788.1    GI:1575796
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 25383)
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
REFERENCE     2 (bases 1 to 25383)
AUTHORS       Waterston,R.
TITLE         The sequence of Homo sapiens cosmid clone U169D2
JOURNAL       Unpublished (1999)
REFERENCE     3 (bases 1 to 25383)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (29-SEP-1996)
REFERENCE     4 (bases 1 to 25383)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (19-JAN-1999) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE     5 (bases 1 to 25383)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (20-JAN-1999) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE     6 (bases 1 to 25383)
AUTHORS       Waterston,R.
```

```
TITLE       Direct Submission
JOURNAL     Submitted (27-APR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY: WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE: This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was mapped by Grieff, M., Whyte, M. P., Thakker, R. V.,
            and Mazzarella, R.  Sequence analysis of 139 kb in Xp22.1
            containing spermine synthase and the 5' region of PEX.  Genomics
            44:227-231 (1997).

            SOURCE INFORMATION:
            This clone is from a chromosome X-specific cosmid library LL0XNCC01
            'U'. The source of the chromosomes was a human/hamster hybrid,
            GM07297-F, from Robert Nussbaum at the University of Pennsylvania
            School of Medicine.  Please contact the Lawrence Livermore National
            Laboratory at http://www.bio.llnl.gov/genome to obtain the clone.
            VECTOR: Lawrist16.
FEATURES             Location/Qualifiers
     source          1..25383
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp22.1-22.2"
                     /clone_lib="LL0XNCC01-U"
                     /clone="U169D2"
     repeat_region   236..527
                     /rpt_family="Alu"
     repeat_region   818..894
                     /rpt_family="L1"
     repeat_region   896..958
                     /rpt_family="L1"
     repeat_region   1122..1429
                     /rpt_family="Alu"
     gene            <1960..16146
                     /gene="H_169D2.1"
     CDS             join(<1960..2076,2989..3169,11533..11627,11988..12183,
                     14388..14463,15924..16146)
                     /gene="H_169D2.1"
                     /note="match to AF019612 (PID:g2745733)"
                     /codon_start=1
                     /product="SP2 metalloprotease"
                     /protein_id="AAD08631.1"
                     /db_xref="GI:4164134"
                     /translation="IWHNFVLALGLGLALVLLPVILLPFYYTGVGVLITEVAEDSPAI
                     GPRGLFVGDLVTHLQDCPVTNVQDWNECLDTIAYEPQYGYCISASTLQQLSFPVRAYK
                     RLDGSTECCNNHSLTDVCFSYRNNFNKRLHTCLPARKAVEATQVCRTNKDCKKSSSSS
                     FCIIPSLETHTRLIKVKHPPQIDMLYVGHPLHLHYTVSITSFIPRFNFLSIDLPVVVE
                     TFVKYLISLSGLALAIVNAVPCFALDGQWILNSFLDATLTSVIGDNDVKDLIGFFILLG
                     GSVLLAANVTLGLWMVTAR"
     repeat_region   3451..3728
                     /rpt_family="Alu"
     repeat_region   6198..6297
                     /rpt_family="HY3"
     repeat_region   7512..7815
                     /rpt_family="Alu"
     repeat_region   8372..8422
                     /rpt_family="MIR"
     repeat_region   9779..10074
                     /rpt_family="Alu"
     repeat_region   10552..10942
                     /rpt_family="Alu"
     repeat_region   12771..13070
                     /rpt_family="Alu"
     repeat_region   13071..13429
                     /rpt_family="L1"
     repeat_region   13442..13744
                     /rpt_family="Alu"
     repeat_region   15145..15446
                     /rpt_family="Alu"
     misc_feature    17715..17839
                     /note="match to EST AA090979 (NID:g1635595)"
     misc_feature    18133..18160
                     /note="match to EST R53421 (NID:g815323) yj70g06.r1"
     repeat_region   18161..18266
                     /rpt_family="L2"
     misc_feature    18267..18453
                     /note="match to EST R53421 (NID:g815323) yj70g06.r1"
     misc_feature    18267..18364
                     /note="match to EST AA090402 (NID:g1635010)"
     misc_feature    complement(18608..18913)
                     /note="match to EST R68268 (NID:g841785) y106c02.s1"
     repeat_region   20528..20828
                     /rpt_family="L1"
     repeat_region   20831..21014
                     /rpt_family="MER1_type"
     repeat_region   21078..21394
                     /rpt_family="L1"
     repeat_region   22041..22216
                     /rpt_family="Alu"
     repeat_region   22218..22318
                     /rpt_family="L2"
     repeat_region   22323..22447
                     /rpt_family="(TA)n"
     repeat_region   22894..23199
                     /rpt_family="Alu"
     repeat_region   24852..24933
                     /rpt_family="L2"
     repeat_region   25152..25383
                     /rpt_family="Alu"
BASE COUNT     7577 a   4608 c   4646 g   8552 t
ORIGIN

alignment_scores:
                  Quality:  38.00                  Length:    12
                    Ratio:  4.222                    Gaps:     0
       Percent Similarity: 75.000       Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x U72788

Align seg 1/1  to: U72788  from: 1  to: 25383

    1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
      ||||||||||||     |||::: |||
24246 TATAGACTAGCAATCTGGAGATTACCCTTAACATAT 24281

seq_name: gb_bal:U00020

seq_documentation_block:
LOCUS       U00020      36947 bp    DNA             BCT       01-MAR-1994
DEFINITION  Mycobacterium leprae cosmid B229.
ACCESSION   U00020
VERSION     U00020.1  GI:467102
KEYWORDS    .
SOURCE      Mycobacterium leprae.
```

```
ORGANISM   Mycobacterium leprae
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Corynebacterineae; Mycobacteriaceae;
           Mycobacterium.
REFERENCE  1 (bases 1622 to 4514)
AUTHORS    Rinke de Wit,T.F., Bekelie,S., Osland,A., Miko,T.L.,
           Hermans,P.W.M., van Soolingen,D., Drijfhout,J.W., Schoeningh,R.,
           Janson,A.A.M. and Thole,E.R.
TITLE      Mycobacteria contain two GroEL genes. The second Mycobacterium
           leprae GroEL gene is arranged in an operon with GroES
JOURNAL    Mol. Microbiol. 6, 1995-2007 (1992)
MEDLINE    92374850
REFERENCE  2 (bases 1 to 36947)
AUTHORS    Smith,D.R.
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 36947)
AUTHORS    Robison,K.
TITLE      Direct Submission
JOURNAL    Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
           School, 200 Longwood Avenue, Boston MA 02115
REFERENCE  4 (bases 1 to 36947)
AUTHORS    Robison,K.
TITLE      Direct Submission
JOURNAL    Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
           School, 200 Longwood Avenue, Boston MA 02115
COMMENT    On Mar 31, 1994 this sequence version replaced gi:414226.
           This sequence data was produced by the Genome Sequencing Center
           located at Collaborative Research Incorporated (1365 Main St.,
           Waltham MA, 02159). 617-487-7979). Please contact Doug Smith
           (smith@cric.com).  The annotation should be considered
           preliminary and incomplete.
FEATURES   Location/Qualifiers
     source          1..36947
                     /organism="Mycobacterium leprae"
                     /db_xref="taxon:1769"
     CDS             complement(1..791)
                     /codon_start=1
                     /transl_table=11
                     /product="B229_C1_175"
                     /protein_id="AAA17292.1"
                     /db_xref="GI:467110"
                     /translation="MTSVSIVVEIGHTSAAEPMLAAAAFGNQPGRWPLPTATTPHQLW
                     LRAVAAGGQGHYSSAYRDLAVLRRSVPAGRLASLAHSTHGSFLRQLGWHSLARGWDGR
                     ALVLAGTDSPARADALIGLAADALGVGRLAAAATLLRRVGSALAPAQLPAQVADRLAV
                     RRRWVAAELAMAVGDGATAVRHAREAVELAQVGRVVSVRHQVKSDVVLAAALCSAATE
                     RARVVAEAALAATGRLGLIPLRWALACLLIDIGSVTFSEPELSELP"
     CDS             1152..1460
                     /note="Probable transcription factor (similarity to
                     Streptomyces coelicolor whiB); B229_F1_2"
                     /codon_start=1
                     /transl_table=11
                     /product="whiB"
                     /protein_id="AAA17313.1"
                     /db_xref="GI:467131"
                     /translation="MPQPKQLPGPNATIWNWQLOGLCRGVDSSMFFHPDGERGARMQ
                     REQRAKEMCRRCPVIEECRAHALDVGEPYCVWGGLSESERDLLLKGDLARSRSIPRSA
                     "
     CDS             complement(2059..3672)
                     /note="GroE1 protein; B229_C3_248"
                     /codon_start=1
                     /transl_table=11
                     /product="groE1"
                     /protein_id="AAA17312.1"
                     /db_xref="GI:467130"
                     /translation="MSKLIEYDETARHAMEVGMNKLADTVRVTLGPRGRHVVLAKAFG
                     GPTINDGVTVAREIDLEDPFENLGAQLVKSVATKNDVAGDGTTATVLAQALVKGG
                     GEGMNKVGTDGVVSVEESSLDTELEFTEGVGFDKGFLSAYFVTDFDSQAVLDDPLV
                     LRMVAAGANPVALGAGISKRAADAVSEALLAVATPVAGKDAITQVATVSSRDEQIGALV
                     LHQEKISSLPELLPMLEKVTESGKPLLIVAEDLEGEALATLVVNSIRKTLKAVAVKS
                     PFFGDRRKAFLEDLAIVTGGQVVNPETGLVLREVGTDVLGSARRVVVSKDDTIVDGG
                     GSNDAVAKRVNQLRAEIEVSDSEWDREKLQERVAKLAGGVAVIKVGAVTETALKKRKE
                     SVEDAVAAAKASIEEGIIAGGGSALVQCGAALKQLRTSLTGDEALGIDVFFEALKAPL
                     YWIATNAGLDGAVVVDKVSGLPAGHGLNASTLGYGDLVADGVVDPVKVTRSAVLNAAS
                     VARMLITETAVDKRPAKTEHDHHGHAH"
     CDS             complement(3751..4053)
                     /note="10 kd chaperonin; B229_C3_247"
                     /codon_start=1
                     /transl_table=11
                     /product="chpA"
                     /protein_id="AAA17311.1"
                     /db_xref="GI:467129"
                     /translation="MAKVKIKPLEDKILVQGEAETMTPSGLVIPENAKEKPQEGTVV
                     AVGPGRWDEDGAKRIPVDVSSEGDIVIYSKIGGTEIKYNGEEYLILSARDVLAVVSK"
     CDS             complement(4517..5389)
                     /note="Probable glycoproteinase (similarity to Pasteurella
                     hemolytica glycoproteinase A1 and E.coli ygjC);
                     B229_C3_246"
                     /codon_start=1
                     /transl_table=11
                     /product="u229e"
                     /protein_id="AAA17310.1"
                     /db_xref="GI:467128"
                     /translation="MTISAVPGTIILAIETSCDETGVGIACLDDYGTVTLLADEVASS
                     VDEQARFGGVVPEIASRAHLEALGPTIRCALAAAGLTGSAKPDVVAATIGPGLAGALL
                     VGVAAAKAYSAAWGVPFYAVNHLGGHLAADVYEHGPLPECVALLVSGGHTHLLQVRSL
                     GAPIVELGSTVDDAAGEAYDKVARLLGLYPGASVHAGDKVARLLGDLLARTGDAIVFPRGMTGPA
                     DDLNAFSFSGLKTAVARYVESHPDALPADVVAAGFQEAVADVLTMKAVRAATGLGVSTL
                     LIVGGVAANSRCAS"
     CDS             complement(5386..6465)
                     /EC_number="2.3.1.128"
                     /note="30S Ribosomal protein S18 alanine
                     acetyltransferase; B229_C1_170"
                     /codon_start=1
                     /transl_table=11
                     /product="rim"
                     /protein_id="AAA17291.1"
                     /db_xref="GI:467109"
                     /translation="MSIVLAIDTATAAVTAGIVAFDGHDCFTLAERVTVDAKAHVERL
                     TPNVLVALADEALAMCELDAVVVGCGPGPFTGLRVGMATAAAYGHALGIPVHGVCSLD
                     AIGVRTTGDILVIVDARRHEVYWARYRDGVRIAGPAVGSPTDVDPGTALIVAGSPEHA
                     ALFGLPLCEPIYPTPAGLVAAVPDWSVSPIPLVALYLRRPDAKPITADNEPIVIGTLT
                     PADVVDRCAQLESQFDGDNPWPAAAFDRELANSYNCYGARTADTLVGYAGITRLGHT
                     PPFEYEVHTIAVDPAYRGRVGRRILGELLDFAGSGAIYLEVRTDNETAIALYRSVGF
                     ERIGLRPRYYPASGADAYLMRREAQ"
     CDS             complement(6462..6947)
                     /note="match to 16.9 kd E.coli protein in amiB region
                     (yjeE); B229_C2_205"
                     /codon_start=1
                     /transl_table=11
                     /product="u229f"
                     /protein_id="AAA17300.1"
                     /db_xref="GI:467118"
                     /translation="MAEYSLRSGAVICERVEDTVALGSRLGEQLRAGDVVLSGPLGA
                     GKTVLAKGIAVAMDVDGPVISPTYLARVHLPRLGTPAMHVDVYRLLDHRDADLVG
                     ELDSLDLDTDLAEAVVVMEWGAGLAECLAARHLDIRLERVRYSDVRIATWQWVCSRDR
                     P"
     CDS             complement(6940..8001)
                     /codon_start=1
                     /transl_table=11
                     /product="B229_C1_169"
                     /protein_id="AAA17290.1"
                     /db_xref="GI:467108"
                     /translation="MKWSPVYAGGSPPERTARPKPLTDDEIRRRRNGRPWLAGGTGLVP
                     VATIVGALSLRSIFERDNACRDPYVDRDFEKLGDERRCWVTISGGMALVVREEGAVKA
                     PVAMVFAYGFYLRMDSFHFQRKRFGKRWGPQVRMVFYDHCGHVQSSEVALDTYTIQL
                     GQDLRTVLQTVTPHGMIVLVGHSMEGILKSPALEAVRLTSRASKLMHRGSIASQSLI
                     GPILRAASYSDLRVSRGLDAFSQRIMNDTLIAILVSFLHALEHEETAGLMPLLRVPA
                     LIACGDHDLLITSDERSRGMAAVLPLLALVIVSGASRLALLDKPGAINDGLVRLVNRAV
                     PGKAALRYRRFKERLQRHG"
     CDS             complement(7931..9097)
                     /EC_number="5.1.1.1"
                     /note="Alanine racemase; B229_C3_243"
                     /codon_start=1
                     /transl_table=11
                     /product="alr"
                     /protein_id="AAA17309.1"
```

                     phosphoribosyl-AMP cyclohydrolase; prcA; prcB; proteasome alpha
                     subunit; proteasome beta subunit; protein translocation system;
                     pseudogene; RLEP; sec-independent.
SOURCE   Mycobacterium leprae.
  ORGANISM   Mycobacterium leprae
                     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
                     Actinomycetales; Corynebacterineae; Mycobacteriaceae;
                     Mycobacterium.
REFERENCE   1 (bases 1 to 40245)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE   Use of an ordered cosmid library to deduce the genomic organization
             of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2 (bases 1 to 40245)
  AUTHORS   Hamlin,N. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 40245)
  AUTHORS   James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Mycobacterium leprae sequencing project,
             Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
             CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
             Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,
             Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
             France Requests for cosmids should be sent to Karin Eiglmeier
             (kei@pasteur.fr)
COMMENT   Notes:
             The Sanger Centre is funded to complete the sequence of M. leprae
             by the Heiser Program for Research in Leprosy and Tuberculosis of
             The New York Community Trust.
             Work in Paris is supported by the Heiser Trust, the Association
             Francaise Raoul Follereau and the Groupement de Recherches et des
             Etudes des Genomes (GIP-GREG).
             Details of M. leprae sequencing at the Sanger Centre are available
             on the World Wide Web.
             (URL, http://www.sanger.ac.uk/Projects/)
             CDS are numbered using the following system eg MLCB33.01c. ML (M.
             leprae), cB33 (cosmid name), .01 (first CDS), c (complementary
             strand).
             The more significant matches with motifs in the PROSITE database
             are also included but some of these may be fortuitous. The length
             in codons is given for each CDS.
             Usually the highest scoring match found by fasta -o is given for
             CDS which show significant similarity to other CDS in the database.
             The position of possible ribosome binding site sequences are given
             where these have been used to deduce the initiation codon. All CDS
             over 100 codons have been analysed. Gene prediction is based on
             positional base preference in codons especially where there is an
             increase in the observed/expected third position G + C. CAUTION:
             We may not have predicted the correct initiation codon. Where
             possible we choose an initiation codon (atg, gtg, or ttg) which is
             preceded by an upstream ribosome binding site sequence (optimally
             5-13bp before the initiation codon). If this cannot be identified
             we choose the most upstream initiation codon.
             IMPORTANT: This sequence MAY NOT be the entire insert of the
             sequenced clone. It may be shorter because we only sequence
             overlapping sections once, or longer, because we arrange for a
             small overlap between neighbouring submissions.
FEATURES   Location/Qualifiers
     source   1..40245
                     /organism="Mycobacterium leprae"
                     /db_xref="taxon:1769"
                     /clone="cosmid B2533"
     gene   complement(1..1099)
                     /gene="ansP"
     misc_feature   complement(1..31687)
                     /note="overlap with EMBL:ML017 cosmid B2126 from 1 to
                     31682. There are 16 conflicts between this sequence and
                     ours. In each case our sequence has been checked and is
                     thought to be correct"
     CDS   complement(<1..1099)
                     /gene="ansP"

                     /db_xref="GI:467127"
                     /translation="MAVTPISLRPGVLAEAVVDLGAIDYNVRVLREHAGMAQLMVVLK
                     ADAYGHGATQVALAALARAGAAELGVATVDEALALRADGISAPVLAWLHPPGIDFGPAL
                     LADVQIAVSSVRQLDELLDAVRRTGRTATVTVKADTGLNRNGVVTDQYPAMITALQRA
                     VVEDAVRLRGLMSHLVYADQPDNPSNDVQGKRFAALLAQAHEQGLRFEVAHLSNSSAT
                     MSRPDLAYDLVRPGIAVYGLSPVPSRGDMGLIPAMTVKCAVAMVKSIRAGEGVSYGHD
                     WIAQHDNLALLPVGYADGVFRSLGGRLDVLINGKRRPGVGRICMDQFVVDLGPGPID
                     VAEGDEAILFGPGARGEPTAQDWADLLGTIHYEVVTSLRGRITRTYREAQTVDR"
     CDS   9105..9446
                     /codon_start=1
                     /transl_table=11
                     /product="B229_F3_101"
                     /protein_id="AAAI7320.1"
                     /db_xref="GI:467138"
                     /translation="MLLFRCVSTRCRHFPRTNGACRSTRVDYGAQLYIMHEVVGAVVS
                     AGLELIQGSPESPSQTTMRSTDISVASSGTVIGGHLPTECPEFVEEIELGVPTVSNEQ
                     GGIPPGWPCGV"
     CDS   complement(9627..9785)
                     /EC_number="4.1.1.15"
                     /note="Glutamate decarboxylase isozyme; B229_C2_202"
                     /codon_start=1
                     /transl_table=11
                     /product="gadD"
                     /protein_id="AAAI7299.1"
                     /db_xref="GI:467717"
                     /translation="MGFLAAPSDVDQCKVAINMGSPYPGVGFCVWRGNEYLPEDVVFR
                     VNYLGGDM"
     CDS   complement(9787..9951)
                     /EC_number="4.1.1.15"
                     /note="Glutamate decarboxylase isozyme; B229_C1_167"
                     /codon_start=1
                     /transl_table=11
                     /product="gadC"
                     /protein_id="AAAI7289.1"
                     /db_xref="GI:467107"
                     /translation="MSWPSLSATFTGELEPVAELCVALDKLAADNGLDIAVSVGAASG
                     GSGCLACTRS"
     CDS   complement(10120..10344)
                     /EC_number="4.1.1.15"
                     /note="Glutamate decarboxylase isozyme; B229_C1_166"
                     /codon_start=1
                     /transl_table=11
                     /product="gadB"
                     /protein_id="AAAI7288.1"
                     /db_xref="GI:467106"
                     /translation="MAETFDKICLDKDEYPAAAPSNGAAYSMATGLYSRLGTQLRQPV
                     QYLWGVPLSAPGMGGDAGLPGHKIALVCGS"
     CDS   complement(10680..11858)

alignment_scores:
   Quality: 38.00   Length: 11
   Ratio: 3.800   Gaps: 0
Percent Similarity: 90.909   Percent Identity: 72.727

alignment_block:
US-08-653-294-19 x U00020   ..

Align seg 1/1   to: U00020   from: 1 to: 36947

        2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
          ||| ||||||||||||||||::||::||||||||
    21906 AGGAGCGGCGATCAGCCAGTGGCCATCAGGTAC 21938

seq_name: gb_bal:MLCB2533

seq_documentation_block:
LOCUS   MLCB2533   40245 bp   DNA   BCT   27-AUG-1999
DEFINITION Mycobacterium leprae cosmid B2533.
ACCESSION   AL035310
VERSION   AL035310.1 GI:4200258
KEYWORDS   ansP; ATP-dependent RNA helicase; ATP-phosphoribosyl transferase;
            ansP; ATP-phosphoribosyl transferase; ATP-phosphoribosyl methyl
            ATPase; hisG 5-methyltetrahydrofolate-homocysteine methyl
            transferase; hisI; L-asraragine permease; metH; mttB;

repeat_unit

```
/note="Pfam match to entry PF00324 aa_permeases, Amino
acid permease, score 501.80, E-value 5.1e-147"
1668..2835
```

misc_feature

```
/note="1168 bp perfect direct repeat"
2104
/gene="ansP2"
```

misc_feature

```
/note="conflict: C is CT in EMBL:ML017"
complement(2500..2592)
/gene="ansP2"
/note="PS00218 Amino acid permeases signature"
complement(2903..3856)
/gene="MLCB2533.03c"
```

gene

```
complement(2903..3856)
/gene="MLCB2533.03c"
```

CDS

```
/note="MLCB2533.03c, hypothetical protein, len: 317 aa;
similar to M.tuberculosis hypothetical protein Rv2125
(MTCY261.21) (EMBL:Z97559) (292 aa), fasta scores; opt:
1648 z-score: 2382.2 E(): 0, 84.1% identity in 290 aa
overlap. Also some similarity to M.leprae hypothetical
protein TR:Q49847 (29.7% identity in 279 aa overlap).
Annotated as ORF TR:Q49797, hypothetical protein in
M.leprae cosmid EMBL:U00017"
/codon_start=1
/transl_table=11
/product="hypothetical protein MLCB2533.03c"
/protein_id="CAA22917.1"
/db_xref="GI:4200261"
/db_xref="SPTREMBL:Q49797"
/translation="MPPHRAVTRHASSALKPYADSVTLRDGGPDRGALPELHNTVVV
AAFEGWNDASDAASGALEHLNAVWEADPIVEIDDEAYYDYQVNRPVIRQVDGVTRELV
WPAMRISYCRPPGSDRNVLMHGVEPNMRWRTFCTELLTIADRLNVDTVVILGALLAD
TPHTRPVPVSGAAYSPESARRFGLEETRYEGPTGIAGVFQDACVAARIPAVMFWAAVP
HYVSHPPNPKATVALLRRVEDVLDVEVPLADLPTQAEDWEQAITEIAAEDDELAEYVH
SLEQRGDAEVDVNDALGKIDGDALAAEFERYLRRRPGFGR"
4004..7555
/gene="metH"
4004..7555
/gene="metH"
```

gene

```
4004..7555
/gene="metH"
```

CDS

```
/note="MLCB2533.04, metH, probable
5-methyltetrahydrofolate-homocysteine methyltransferase,
len: 1183 aa; similar to many members of vitamin-B12
dependent methionine synthase family e.g. METH_ECOLI
(EMBL:X16584) E.coli metH (1226 aa), fasta scores; opt:
1617 z-score: 1000.7 E(): 0, 31.6% identity in 1228 aa
overlap. Equivalent to M.tuberculosis Rv2124c
(MTCY261.20c, 88.7% identity in 1183 aa overlap).
Annotated as METH_MYCLE, designated metH2 in M.leprae
```

alignment_scores:
```
          Quality:    38.00             Length:       12
            Ratio:     4.222               Gaps:        0
Percent Similarity:   75.000    Percent Identity:   66.667
```

alignment_block:
US-08-653-294-19 x MLCB2533

Align seg 1/1 to: MLCB2533   from: 1   to: 40245

```
           1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
             |||||||||||    |||||:::  |||||
       17534 TATCGATTGGCAAGTCGCAGAATATCTGCACGGTAT 17569
```

seq_name: gb_ba1:U00017

```
/note="MLCB2533.01c, ansP, probable L-asparagine permease,
partial CDS, len: >366 aa; highly similar to many
amino-acid permeases e.g. ANSP_SALTY (EMBL:U04851)
S.typhimurium AnsP, L-asparagine permease (L-asparagine
transport protein) (497 aa), fasta scores; opt: 1508
z-score: 1696.8 E(): 0, 61.2% identity in 366 aa overlap.
Equivalent to M.tuberculosis Rv2127, ansP (MTCY261.26,
85.7% identity in 356 aa overlap). Also similar to
M.tuberculosis permease Rv0346c, aroP2 (MTCY13E10.06c,
75.1% identity in 345 aa overlap). Probable integral
membrane protein, contains PS00218 Amino acid permeases
signature. Pfam match to entry PF00324 aa_permeases, amino
acid permease. Annotated as ORF TR:Q49801, designated
aroP2 in M.leprae cosmid EMBL:U00017"
/codon_start=1
/transl_table=11
/product="putative L-asparagine permease"
/protein_id="CAA22915.1"
/db_xref="GI:4200259"
/translation="MATLAESPEPKSGASRAGVLGEEAGYHKGLKPRQLQMIGIGGAI
GTGLFLGAGGRLAKAGPGLFLVVAVCGVFVFLLLRALGELVLHRPSSGSFVSYAREFF
GEKAAYVVGWLYFLDWAMTAIVDTTAIATYLHRWTIFTALPQWTLALLALAVVLVMNL
ISVEWFGELEFWAALIKVCALMAFLVVGTIFLGGRYPVDGHNTGLSLWTSHGGLFPTG
VAQLIVVSSGVMFAYAAVELVGTAAGETVEPKKIMPRAINSVIARIAIFYVGSVILLA
LLLPYSAFKASESPFVTFFSKVGFYGAGDLMNIVVLTAALSSLNAGLYATGRVMHSIA
INGSGPKFTARMSKNGVPYGGILLAAVICLCG"
```

repeat_unit
```
1..1168
/note="1168 bp perfect direct repeat"
```

misc_feature
```
complement(2..991)
/gene="ansP"
```

unsure
```
/note="Pfam match to entry PF00324 aa_permeases, Amino
acid permease, score 245.80, E-value 6e-70"
438
/gene="ansP"
/note="ambiguous base T /G"
```

misc_feature
```
complement(833..925)
/gene="ansP"
/note="PS00218 Amino acid permeases signature"
1223
```

misc_feature
```
/note="conflict: T is TG EMBL:ML017"
complement(1249..2766)
/gene="ansP2"
```

gene
```
complement(1249..2766)
/gene="ansP2"
```

CDS
```
/note="MLCB2533.02c, ansP2, probable L-asparagine
permease, len: 505 aa; highly similar to many amino-acid
permeases e.g. ANSP_SALTY (EMBL:U04851) S.typhimurium
AnsP, L-asparagine permease (L-asparagine transport
protein) (497 aa), fasta scores;opt: 1891 z-score: 2218.9
E(): 0, 58.9% identity in 477 aa overlap. Equivalent to
M.tuberculosis Rv2127 (MTCY261.26, 83.7% identity in 485
aa overlap). Also similar to M.tuberculosis permease
Rv0346c (MTCY13E10.06c, 69.8% identity in 473 aa overlap).
Probable integral membrane protein, contains PS00218 Amino
acid permeases signature. Pfam match to entry PF00324
aa_permeases, amino acid permease. Annotated as ORF
TR:Q49802, designated lysP in M.leprae cosmid EMBL:U00017"
/codon_start=1
/transl_table=11
/product="putative L-asparagine permease"
/protein_id="CAA22916.1"
/db_xref="GI:4200260"
/translation="MATLAESPEPKSGASRAGVLGEEAGYHKGLKPRQLQMIGIGGAI
GTGLFLGAGGRLAKAGPGLFLVVAVCGVFVFLLLRALGELVLHRPSSGSFVSYAREFF
GEKAAYVVGWLYFLDWAMTAIVDTTAIATYLHRWTIFTALPQWTLALLALAVVLVMNL
ISVEWFGELEFWAALIKVCALMAFLVVGTIFLGGRYPVDGHNTGLSLWTSHGGLFPTG
VAPLIVVSSGVMFAYAAVELVGTAAGETVEPKKIMPRAINSVIARIAIFYVGSVILLA
LLLPYSAFKASESPFVTFFSKVGFYGAGDLMNIVVLTAALSSLNAGLYATGRVMHSIA
INGSGPKFTARMSKNGVPYGGILLAAVICLCGVALNAFNPGQAFEIVLSVAALGIIAG
WGTIVLCQLRLHKMAKAGIMRRPFRMPLAPYSGYLTLAFLFAVLVVMAFDKPIGTWT
VASLIVIVPALIAGWYSIRKRVMTIARERMGYTGPFPAIANPVQPSERSHSQNP"
```

misc_feature
```
complement(1354..2691)
/gene="ansP2"
```

```
         Actinomycetales.; Corynebacterineae; Mycobacteriaceae;
         Mycobacterium.
REFERENCE  1  (bases 1 to 42157)
  AUTHORS  Smith,D.R.
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 42157)
  AUTHORS  Robison.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
           School, 200 Longwood Avenue, Boston MA 02115
REFERENCE  3  (bases 1 to 42157)
  AUTHORS  Robison.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
           School, 200 Longwood Avenue, Boston MA 02115
COMMENT    On Mar 31, 1994 this sequence version replaced gi:414223.
           This sequence data was produced by the Genome Sequencing Center
           located at Collaborative Research Incorporated (1365 Main St.,
           Waltham MA, 02159).  617-487-7979).  Please contact Doug Smith
           (smith@cric.com).  The annotation should be considered
           preliminary and incomplete.
FEATURES             Location/Qualifiers
     source          1..42157
                     /organism="Mycobacterium leprae"
                     /db_xref="taxon:1769"
     CDS             complement(133..1086)
                     /note="match to yigU and yigV E.coli; B2126_C1_183"
                     /codon_start=1
                     /transl_table=11
                     /product="u2126a"
                     /protein_id="AAA17191.1"
                     /db_xref="GI:467006"
                     /translation="MNGCRAVVRACDLLRRIKQHYRRSRTNPDATMSLIDHLTELRTR
                     LLISLAAIVVTTIFGFIWYSHSIFGLESLGEWLRRPYCSLPQSARADISPDGQCRLLA
                     TAPFDQFMLRIKVGMAAGIVLASPVWFYQLWAFITPGLYTKERRFTVAFVVPAAVLFA
                     GGTVLAYLVLSKALGFLLIVGSGVQVTALSGDRYFGFLLNLLVFGVSFEFPLLIVML
                     NIAGLLTYQRLKSWRRGLIFAMFVFAAVFTPGSDPFSMTALGAALTVLLELAIQLVRL
                     HDKRRVKHEALIADDEASVIEPPSSIPERSYTAIRSHDDVT"
     CDS             complement(1132..1398)
                     /note="match to yigT E.coli; B2126_C1_182"
                     /codon_start=1
                     /transl_table=11
                     /product="u2126b"
                     /protein_id="AAA17190.1"
                     /db_xref="GI:467005"
                     /translation="MGSLSPWHWVVLVVVVLLFGAKKLPDAARSLGKSMRIFKSELR
                     EMQTENQAQASALETPMQNPTVVQSQRVVPPWSTEQDHTEARPA"
     CDS             complement(1471..1692)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C1_181"
                     /protein_id="AAA17189.1"
                     /db_xref="GI:467004"
                     /translation="MSWIFEYYPMRQAHELPDGSFQAVMTYASDAWMTRLVLGFGSAV
                     QVQAPEALAYRVRNAAVAALESYQVTAQA"
     CDS             complement(1701..2444)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C2_220"
                     /protein_id="AAA17199.1"
                     /db_xref="GI:467014"
                     /translation="MTQLSTRLVRLLNMVPYFQANPRITRAEAAADLGVSAKQLDQDF
                     NQLMVCGLPGYGPGDLIDFEFSGDIEVTFSAGIDRPLQLTSPEAIGLLVALRALANI
                     PGVVDPEAVRSAIARKIEAAAVMGNEATGSVASVDTRPFSESHAVAVAVAVRNKQAL
                     VIDYYSASHDTLTSRIVDPIRVLLVGDHSYLEAWSREAEGVRLFRFDRIVVARELDEP
                     AAAPETVRRRCQTHRSSMTTRCCRRRRCG"
     CDS             complement(2441..3436)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C3_266"
                     /protein_id="AAA17208.1"
                     /db_xref="GI:467023"
                     /translation="MATSKVERLVNLVIALLSTRGYMTAEKIRSSVAGYSDSPTVEAF
```

```
                     SRMFERDKNELRDLGIPLEVGKVSALDPSEGYRINRDAYALPPVELTPDEAAAVAVAT
                     QLWESQELITATQGALLKLRAAGVDIDPLDTPVVIASSGVSSLRGSEDFLSTLLSAI
                     GSRQAVQFPYRPSRAEPYTMRNVEPWGVITENSCWYLVGHDCDRNATRTFRLSRIGSE
                     VAPIGPAGAVTVPDGVDLRRIVSDAVAEVSTGATARWWVDGRATALRHAGRPAGVRR
                     LGGRDGQVIELDIGSIDRLARDIAGHGADAVVLEPDALRDDVLIRLRAHAGTGPS"
     CDS             complement(3443..3841)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C3_265"
                     /protein_id="AAA17207.1"
                     /db_xref="GI:467022"
                     /translation="MDTEIDWVIKRRLFQRYQDRDMELTDPKIAQLDLAYHDIKRGR
                     GVFDLLQRKGLAARVTTDEDIADAVNHPPQTTRARLRGEFISAAQAAGRDFTVDWVHL
                     KLNDQAQRTVLCKDPFRAVDERVKRLIASM"
     CDS             complement(3861..4871)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C2_219"
                     /protein_id="AAA17198.1"
                     /db_xref="GI:467013"
                     /translation="MVGWYPAWLAHRDAQTSTLKDVQRRIMGIETEFGVTCTFHGH
                     RRLSPDEVARYLFRRVVSWGRSSNVFLRNGARLYLDVGSHPEYATAECDNLVQLVTHD
                     RAGEWVLEDLLVDAEQRLADESIGGDIYLFKNNTDSAGNSYGCHENVLIVRAGEFSRI
                     SDVLLPFLVTRQLICGAGKVLQTPKAATFCLSQRAEHIWEGVSSATTRSRPIINTRDE
                     PHADAEKYRRLHVIVGDSNMCETTMLKVTAALMELMEVGVSFPRDFSLDNPIRAIR
                     EVSHDITGRRPVRLAGGRQASALDIQREYYTRAFEHLQTREPNVQFEQVVDLWGPSAR
                     RH"
     CDS             complement(6569..6862)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C3_261"
                     /protein_id="AAA17206.1"
                     /db_xref="GI:467021"
                     /translation="MKIADANVLLYSGNTSSEHTGHPCAGSTVRLSCADRIGFGWVPL
                     LVFVRPATKMGLVLRTMSSEDAIGQVADWLTGPSAVLMCLTVRHAAFLVKIIV"
     CDS             complement(6822..7115)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C2_217"
                     /protein_id="AAA17197.1"
                     /db_xref="GI:467012"
                     /translation="MSFKKVFNDVIRDSVLGRFEPHRFTACITDLALSAVNLDWALPI
                     AAEFEDEELVRRLSGPGHERRSFLSHRLAPHLPARIITIHENRRRERAAVFR"
     CDS             complement(7253..8050)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C3_250"
                     /protein_id="AAA17205.1"
                     /db_xref="GI:467020"
                     /translation="MSFPYFISPEQAMREQRELARKGIARGRSVVALAYAGGVLFVAE
                     NPSRSLQKISELYDRVGFAAAGKFNEFDNLRRGGIQFADTRGYAYDRRDVTGRQLANV
                     YAQTLGTIFTEQAKPYEVELCVAEVAHYGETKPPELYRITIYDGSINDEPHFMVMGGTT
                     ESIANALKESYAENASLTDALGIAVAALRAGSADAAGSDQPTLGVASLEVAVLDANRP
                     RRAFRRIIGSGLEALLREKDSKGSKGAQNPKGARDSKNSKSYGESTD"
     CDS             complement(8047..8922)
                     /note="proteasome, beta subunit: B2126_C1_173"
                     /codon_start=1
                     /transl_table=11
                     /product="prcB"
                     /protein_id="AAA17188.1"
                     /db_xref="GI:467003"
                     /translation="MTRSFPDRLPTNLAFPGISVINQSSFVDLLRRQAPELPVSLGG
                     GSAGGGGQQLSGTTIVVLKYPGGVVIAGDRRSTQGNMIAGRDVRKYVITDDYTATGIA
                     GIAAVVEFARLVAAVMGNEATGSVASVDTRPFSESHAVAVAVAVRNLTAAMGGLLALPLL
                     AGYDIHAPDPQSAGRIVSFDAAGGWNIEEEGYQSVGSGSIFAKSSIKKLYSQVSDADS
                     ALRVAIEALYDAADDSATGCPDLVRGIYPTAVTICAEGAEVTESRIAELAREIIES
                     RSRAYTLGSFGGSEK"
     CDS             complement(9142..10686)
                     /codon_start=1
                     /transl_table=11
                     /product="B2126_C1_172"
                     /protein_id="AAA17187.1"
                     /db_xref="GI:467002"
```

```
                    euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
                    eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
                    Arabidopsis.
REFERENCE   1  (bases 1 to 103125)
AUTHORS     Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
            Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
TITLE       Arabidopsis thaliana chromosome II BAC F2I9 genomic sequence
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 103125)
AUTHORS     Rounsley,S.D. and Lin,X.
TITLE       Direct Submission
JOURNAL     Submitted (29-AUG-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 103125)
AUTHORS     Rounsley,S.D.
TITLE       Direct Submission
JOURNAL     Submitted (23-OCT-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Oct 23, 1998 this sequence version replaced gi:3776582.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone F2I9 is from Arabidopsis chromosome II and is near the
            molecular marker rga.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES             Location/Qualifiers
     source          1..103125
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="rga"
                     /clone="F2I9"
     mRNA            <41..>697
                     /gene="F2I9.1"
     gene            <41..>697
                     /gene="F2I9.1"
                     /note="predicted by genscan"
     CDS             41..697
                     /gene="F2I9.1"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAC67316.1"
                     /db_xref="GI:3785969"
                     /translation="MKQNLVAVDILLNQMRDRNIKLDSFVVSAIIETNCDRCRTEGAS
                     LAPDYSLEMGIHLKKSAYLALIGNFLRSNELPKVIEVVKEMVKAQHSLGCYQGAMLIH
                     RLGFGRRPRLAADVFDLLPDDQKGVAAYTALMDVYISAGSPEKAMKILREMREREIMP
```

```
                    /translation="MQRIGTEVEYGISSPSDPTANPILTSTQAVLAYAAAAGIQRVK
                    RTRWDYEVESPLRDARGFDLSRSAGPPPVDADEVGAANMILTNGARLYVDHAHPEYS
                    APECTDPLDAVIWDKAGERVMEAAARHVASVPGAAKLQLYKNNVDGKGASYGAHENYL
                    MSRQTPFSAIIAGLTPFLVSRQVVTGGSRVGIGPAGDEPGFQLSQRSDYIEVEVGLET
                    TLKRGIINTRDEPHADADRYRRLHVIVGDANLAETSTYLKLGTALVLDLIEGPVHG
                    IDLTDLTLARPVHAVHAISRDASLRARTVTLVDGRELTGLALQRIYLDRVAKLVDSRDP
                    DRRAADVVKTWVHVLDQLERDPMDCAELLDWPAKLRLLEGFRQRENLNWSAPRLHLVD
                    LQYSDVRLDKGLYNRLVARGSMKRLVNEHQVLRAVNNPPTDTRAYFRGECLRRFSADI
                    AAASWDSVIFDLGGDSLVRIPTLEPLRGSKAHVGALLDSVDSAAELVEQLTTKPVDPG
                    KVGLDR"
repeat_region   11019..11611
                    /rpt_family="RLEP"
CDS             complement(11986..12465)
                    /codon_start=1
                    /transl_table=11
                    /product="B2126_C1_169"
                    /protein_id="AAA17186.1"
                    /db_xref="GI:467001"
                    /translation="MSSAILYLYARLFLLWMCKEEVVAELDDSFYADYLDLNCKYPDV
                    QMFFSVLPNIFVAARPSAKTAVIAVIILKGPFHFTGGAPALDGAKSTFKHDDEWTDEW
                    AHGRSLDALTLQPRHRLQFNIRISDLECEDSRIAALISTDSRRLACDYSLPALSDAA"
CDS             complement(12950..13234)
                    /codon_start=1
                    /transl_table=11
                    /product="B2126_C3_259"
                    /protein_id="AAA17204.1"
                    /db_xref="GI:467019"
                    /translation="MAVGDDDDRHSAGLLSSIDRHDWGWLQPQSSPAASTVQPLPASR
                    IPKTSFKPYHVPMLGSRGKASSSTLAALVQAALDPVKQNCTRSRVLFGT"
CDS             13405..13665
                    /codon_start=1
                    /transl_table=11
                    /product="B2126_F1_17"
                    /protein_id="AAA17209.1"
                    /db_xref="GI:467024"
                    /translation="MVIVKSERVADLLSSVLVVARNTDLGYQRFGIRAGMGDAHDTVT
                    PSDSRCPGGPNPPVVLTTLIVFNTYWIVATTLIREHYGWR"
CDS             complement(13640..13921)
                    /codon_start=1
                    /transl_table=11
                    /product="B2126_C3_258"
                    /protein_id="AAA17203.1"
                    /db_xref="GI:467018"
                    /translation="MVAVTTATIYICISCEVAITKTTLRRSALVSAALGEASSIAATG
                    SVRGPAFRSSYGSSTVDFIVGLTRFDGGGEDTGAFAVPVRKIIASHSAP"
alignment_scores:
    Quality:          38.00           Length:          12
    Ratio:             4.222          Gaps:             0
Percent Similarity:  75.000     Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x U00017/rev    ..

Align seg 1/1  to reverse of: U00017  from: 1  to: 42157

     1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
       ||||||||||  |||||||||::::   ||||||
 14153 TATCGATTGGCAAGTCGCAGAATATCTGCACGGTAT 14118

seq_name: gb_pl2:ATAC005560

seq_documentation_block:
LOCUS       ATAC005560  103125 bp    DNA            PLN       23-OCT-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC F2I9 genomic sequence,
            complete sequence.
ACCESSION   AC005560
VERSION     AC005560.1  GI:3785968
KEYWORDS    HTG.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
                SLGTYDVLLSGLEKTSDFQKEVALLRKEKKSLVASARFRENVHVEDKICDLLFATNLL
                "
repeat_region   complement(785..827)
                /rpt_family="(TAAAA)n"
mRNA            join(<875..994,1320..>1640)
                /gene="F2I9.2"
gene            <875..>1640
                /gene="F2I9.2"
CDS             join(875..994,1320..1640)
                /gene="F2I9.2"
                /note="predicted by genscan"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAC67317.1"
                /db_xref="GI:3785970"
                /translation="MARSCSIWTPVLISLSPVTGESPKISRRVILATSIGSPQPLLE
                AKIKEPQKLQVLDSKDVSRRNTMLYLAAGFLGGINFLNGEAAEARVGRKENRKKALEK
                LRAKAKESEPNNKSGNQKIEKELEKEEVFPLLPPPLVVEANLLQ"
misc_feature    complement(1683..1725)
                /note="exon predicted by xgrail, quality marginal"
mRNA            complement(<1769..>2932)
                /gene="F2I9.3"
gene            complement(<1769..>2932)
                /gene="F2I9.3"
                /note="predicted by genefinder"
CDS             complement(1769..2932)
                /gene="F2I9.3"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAC67318.1"
                /db_xref="GI:3785971"
                /translation="MKANMSSSSSTRYRSFVFPILFAVLFLGLLIPSSADNRHVINF
                RSPGLYPEGLTWDPRDQHFLVGSLHSRTIHSVSDAGVIETLISDLDLPENSTILGLAV
                DSTNRRLLACIQSLPPLPPFSALASYDLRSGGRRVFLSPLPSLPGDDEDIARDVANDV
                AVDFKGNAYVTNSAKNFIWKVDRDGAASIFSKSPLFNSQPVAADADASFRDCGLNGIV
                YISKGYLLVVQSNTGKVFKVDEDSGNARLVLNGDLIAADGMTRRRDGTVMVVSQKK
                LWLLKSQDSWSEGVVYDEIDLDIEGFPTAVTVAGDRIYVLYGRVMEGIMGSSYKEEG
                AREWFGIEEVVSEKEGGEDKIWLYLIGFGFAYFCFWRFOMKKLITNMDKKIT"
misc_feature    complement(3756..3826)
                /note="exon predicted by xgrail, quality good"
misc_feature    complement(4609..4669)
                /note="exon predicted by xgrail, quality marginal"
mRNA            complement(join(6667..7104,7209..7285,7380..7537,
                7915..8000,8140..8407,8565..>9759))
                /gene="F2I9.4"
gene            complement(6667..>9759)
                /gene="F2I9.4"
CDS             complement(join(7038..7104,7209..7285,7380..7537,
                7915..8000,8140..8407,8565..9759))
                /gene="F2I9.4"
                /codon_start=1
                /product="putative auxin transport protein"
                /protein_id="AAC67319.1"
                /db_xref="GI:3785972"
                /translation="MITWHDLYTVLTAVPLYAMILAYGSVQWWKIFSPDQCSGINR
                FVAIFAVPLLSFHFISTNDPYAMNFRFVAADTLQKIIMLVLLALWANLTKNGSLEWMI
                IMEQFPETGASIVSFKVESDVVSLDGHDFLETDAEIGNDGKLHVTVRKSNASRRSLMM
                TPRPSNLTGAEIYSLSSTPRGSNFNHSDFYSVMGFPGGRLSNFDPADLYSVQSSRGPT
                PRPSNFEENNAVKKYGFYNNTNSSVPAAGSYPAPNPEFSTGTGVSTKPNKIPKENQQL
                QEKDSKASHDAKELMFVWSSSASPVSDVFGGGAGDNVATEQSEQGAKEIRMVVSDQP
                RKSNARGGDDIGGLDSGEGEREIFKATAGLNKMGSNSTAELEAAGGDGGGNNGTHMP
                PTSVMTRLLIMVRKLIRNPNTVYSSLIGLIWALVAYRWHVAMPKILQOSISILSDAG
                LGMAMFSLGLFMALQPKIIACGNSVATFAMAVRFITGPAIMAVAGIAIGLHGDLLRIA
                IVQAALPGGIVPFVFAKEYNVHPTILSTGVIFGMLIALPITLVYYILLGL"
misc_feature    10020..10129
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   complement(10403..10520)
                /rpt_family="AT_rich"
misc_feature    10586..10649
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
```

```
misc_feature    10803..10939
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
misc_feature    complement(11141..11230)
                /note="exon predicted by xgrail, quality excellent"
repeat_region   11408..11446
                /rpt_family="AT_rich"
misc_feature    11620..11654
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(12452..12519)
                /note="exon predicted by xgrail, quality excellent"
misc_feature    12587..12619
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(13191..13230)
                /note="exon predicted by xgrail, quality good"
misc_feature    complement(13329..13500)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    13535..13672
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   13852..13964
                /rpt_family="POLY_A"
misc_feature    complement(13972..14049)
                /note="exon predicted by xgrail, quality marginal"
mRNA            complement(join(<14358..14616,15958..16037,16158..>16307))
                /gene="F2I9.5"
gene            complement(<14358..>16307)
                /gene="F2I9.5"
CDS             complement(join(14358..14616,15958..16037,16158..16307))
                /gene="F2I9.5"
                /codon_start=1
                /product="putative homeobox protein"
                /protein_id="AAC67320.1"
                /db_xref="GI:3785973"
                /translation="MNRLPSSEDGDDEEFSHDDGSAPPRKKLRLTREQSRLLEDSFRQ
                NHTLNPKQKEVLAKHLMLRPRQIEVWFQNRRARSKLKQTEMECEYLKRWFGSLTEENH
                RLHREVEELRAMKVGPTTVNSASSLTMCPRCERVTPAASPSRAVVPVPAKKTFPPQER
                DR"
repeat_region   14881..14924
                /rpt_family="AT_rich"
misc_feature    complement(16632..16710)
                /note="exon predicted by xgrail, quality excellent"
repeat_region   complement(16998..17033)
```

```
alignment_scores:
                Quality:      38.00         Length:       11
                Ratio:         3.455         Gaps:          0
Percent Similarity: 100.000   Percent Identity: 72.727

alignment_block:
US-08-653-294-19 x ATAC005560/rev

Align seg 1/1  to reverse of: ATAC005560   from:   1   to: 103125

            2 ArgLeuAlaLeu***ArgIleAlaLeuArgTyr 12
              ::::|||::||||:::||:::|||||||||||
        44994 AAGTGGGAATTACTAGAGTTGCATTAAGATAT 44962

seq_name: gb_htg1:CNS01DTM

seq_documentation_block:
LOCUS       CNS01DTM    191442 bp    DNA                  HTG       04-NOV-1999
DEFINITION  Homo sapiens chromosome 14 clone R-80A15, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL132800
VERSION     AL132800.1   GI:6272127
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 191442)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continue. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage, etc.
            Contig order : 13 10 12 11 14 15 9, 1000 N's separate segments
            Contig 13 : length 15976 bp
            Contig 10 : length 5980 bp
            Contig 12 : length 9512 bp
            Contig 11 : length 8124 bp
            Contig 14 : length 67723 bp
            Contig 15 : length 81177 bp
            Contig 9 : length 4117 bp.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..191442
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone_lib="RPCI-11"
                     /clone="R-80A15"
     STS             67422..67594
                     /standard_name="M69137"
                     /note="matching EMBL:M69137; Identified using the e-PCR
                     software (G. Schuler)"
     STS             132481..132607
                     /standard_name="AA194835"
                     /note="matching EMBL:AA194835; Identified using the e-PCR
                     software (G. Schuler)"
     STS             133712..133887
                     /standard_name="z41390"
                     /note="matching EMBL:z41390; Identified using the e-PCR
                     software (G. Schuler)"
     STS             133741..133892
                     /standard_name="R39150"
                     /note="matching EMBL:R39150; Identified using the e-PCR
                     software (G. Schuler)"
     STS             133768..133909
                     /standard_name="T31019"
                     /note="matching EMBL:T31019; Identified using the e-PCR
                     software (G. Schuler)"
     STS             146749..146899
                     /standard_name="R37382"
                     /note="matching EMBL:R37382; Identified using the e-PCR
                     software (G. Schuler)"
     STS             146875..146986
                     /standard_name="R52665"
                     /note="matching EMBL:R52665; Identified using the e-PCR
                     software (G. Schuler)"
     STS             154585..154748
                     /standard_name="T15974"
                     /note="matching EMBL:T15974; Identified using the e-PCR
                     software (G. Schuler)"
BASE COUNT   53877 a  41528 c  39774 g  50165 t  6098 others
ORIGIN

alignment_scores:
        Quality:   38.00          Length:      12
        Ratio:     3.800          Gaps:         0
Percent Similarity: 83.333    Percent Identity: 50.000

alignment_block:
```

```
US-08-653-294-19 x CNS01DTM

Align seg 1/1  to: CNS01DTM  from: 1  to: 191442

        1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
          |||||||||   ::   ||::::::::||||||
     1830 TACAGACTTCTTTGATTAGAGTGTGCATGAGATAT 1865

seq_name: gb_in1:DVU57315

seq_documentation_block:
LOCUS       DVU57315        1911 bp    DNA              INV       21-MAR-1997
DEFINITION  Drosophila virilis LAMA (lama) gene, partial cds.
ACCESSION   U57315
VERSION     U57315.1  GI:1314865
KEYWORDS    .
SOURCE      Drosophila virilis.
  ORGANISM  Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1911)
AUTHORS     Perez,S.E. and Steller,H.
TITLE       Molecular and genetic analyses of lama, an evolutionarily conserved
            gene expressed in the precursors of the Drosophila first optic
            ganglion
JOURNAL     Mech. Dev. 59 (1), 11-27 (1996)
MEDLINE     97047309
REFERENCE   2  (bases 1 to 1911)
AUTHORS     Perez,S.E. and Steller,H.
TITLE       Direct Submission
JOURNAL     Submitted (30-APR-1996) Sharon E. Perez, Biology, MIT, 77
            Massachusetts Ave., Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..1911
                     /organism="Drosophila virilis"
                     /db_xref="taxon:7244"
     gene            1..1911
                     /gene="Dvir\lama"
                     /allele=""
                     /db_xref="FlyBase:FBgn0015684"
     CDS             <1..1911
                     /gene="Dvir\lama"
                     /note="lamina ancestor"
                     /codon_start=1
                     /product="LAMA"
                     /protein_id="AAB49927.1"
                     /db_xref="GI:1314866"
                     /translation="WAQIEVETQRSYPDWVQAYAAGILEGSLTWKNIYNQWAKKYEFS
                     SLATHSYSNFLPALSTISSCDRDESSOKFCECVRNLLTRNYEQLKEQAAAMAEHDH
                     YWHQLHILNQLEGMETGYIRGATRARSDLEEEIPLSDFLLMNAAADIQDLKIYYENY
                     VLPNGSAECGRQAGSKNFFLPSASMLTRLLRDEQAPQQSLQLLFGHSTAGSYSSMLRI
                     QKRYKFHYHFSPDDYRSNTVPGADTTFTGYPGIIGSTDDFYVVKGRQVQSIVGGVCIKN
                     ENLALWEKVDVKNMVPLVARVMAANRIAQNRRTWARAMSRHPFTGAKQWISVDLNKLG
                     AQDNLYNTLDADEKHDDAPVPLNEKDNAAISERHDQLKNMWIVEQLPGLVHSKDVTE
                     NFLLAGNSTWLANGVPYFDVILNASRISRDNYSEDQDLTPAEEAVPYNYSEDQDLTPA
                     EEAELTNLEAVDKYLRNGGFRGDLLGDESIAYGNIDLKCYFSYNARLGMSDYHAFAGP
                     IFLRLQHAQARSSLEEPQLDQADQAAPAIGDERLSVIDDAHTLAELELITERRPVR
                     NDMRAIAMRKIGSGPFKWSAMSVLDDGNHAGHPDEWNFDKVSPRWVSVSPIVLCYVPH
                     RISPATTATEKQ"
BASE COUNT  424 a  536 c  567 g  384 t
ORIGIN

alignment_scores:
        Quality:   37.00          Length:      11
        Ratio:     3.364          Gaps:         0
Percent Similarity: 100.000   Percent Identity: 72.727

alignment_block:
US-08-653-294-19 x DVU57315/rev

Align seg 1/1 to reverse of: DVU57315  from: 1  to: 1911
```

```
   1 TyrArgLeuAlaIle***ArgIleAlaLeuArg 11
     :::||::|||||||||||||||::::||||||
1727 TTCGCATGGCTATCGCACGCATATCATTACGGC 1695

seq_name: gb_pr3:HS782L23

seq_documentation_block:
LOCUS       HS782L23   123925 bp   DNA           PRI   23-NOV-1999
DEFINITION  Human DNA sequence from clone 782L23 on chromosome 1p31.2-33
            Contains start of HOOK1 gene, ESTs, STS and GSSs, complete
            sequence.
ACCESSION   AL035416
VERSION     AL035416.7  GI:4775629
KEYWORDS    HTG; CpG Island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123925)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 11, 1999 this sequence version replaced gi:4741514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the overlapping clone
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL
            This sequence is the entire insert of clone 782L23. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            782L23 is from the library RPCI4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..123925
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p31.2-33"
                     /clone="RP4-782L23"
                     /clone_lib="RPCI-4"
     mRNA            complement(join(26. .108,605. .639,2393. .2486,4942. .5082,
                     6225. .6285,6755. .6842,16087. .16197,16754. .16955,
                     18085. .18225,21672. .21838,24879. .24962,28335. .28397,
                     29858. .29925,31733. .31865,33057. .33107,36418. .36490,
                     43327. .43412,50090. .50318))
                     /gene="dJ782L23.1"
                     /note="match: cDNAs: Em:AF044923; match: ESTs:
                     Em:AA161507"
                     /evidence=not_experimental
                     /product="dJ782L23.1 (HOOK1)"
     gene            complement(26. .50318)
                     /gene="dJ782L23.1"
     CDS             complement(join(<26. .108,605. .639,2393. .2486,4942. .5082,
                     6225. .6285,6755. .6842,16087. .16197,16754. .16955,
                     18085. .18225,21672. .21838,24879. .24962,28335. .28397,
                     29858. .29925,31733. .31865,33057. .33107,36418. .36490,
```

```
                     43327. .43412,50090. .50152))
                     /gene="dJ782L23.1"
                     /note="match: proteins: Tr:060561"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ782L23.1 (HOOK1)"
                     /protein_id="CAB52263.1"
                     /db_xref="GI:5706448"
                     /translation="MEETQPPPOPKLPLCDSLIMWLQFNTASPCQDVKQLTSGVAMA
                     QVLHQIDAAWFNESWLSRIKEDVGDNWRIKASNVKKVLQGIMSYYHEFLGQISEALI
                     PDLNQITECSDPVELGRLLQLILGCAINCEKKQEHIQNIMTLEESVQHVVMTAIQELM
                     SKEILSSPPNDAVGELEQQLKRALEELQEALAKEELRQRCEELDMQVTTLQDEKNSL
                     VSENEMNEKLDQLDGSFDDPNTVVAKKYFHAQLQELQEENFRLEAKDDYRVHCE
                     ELEKQLIEFQHRNDELTSLAEETRALKDEIDVLRATSDKANKLESTVEIYRQKLQDLN
                     DLRKQVKTLQETNMYMHNTVSLEEELKANAARTQLETYKROVQDLHVKLSSESKRA
                     DTLAFEMKRLEEKHEALLKEKERLIEQRDTLKETNEELRCSQVQQDHLNQTDASATKS
                     YENLAAEIMPVEYREVFIRQHENKMLRLQQEGSENERIEELQEQLEQKHRKMNELET
                     EQRLSKERIRELQQIEDLQKSLQEOGSKSEGESSSLKQKLEAHMEKLTEVHEELQK
                     KQELEDLQPDINQN"
     misc_feature    complement(<28. .638)
                     /gene="dJ782L23.1"
                     /note="match: ESTs H90470 T78548 T85713"
     repeat_region   134. .253
                     /note="MIR repeat: matches 61. .188 of consensus"
     misc_feature    complement(<601. .5013)
                     /gene="dJ782L23.1"
                     /note="match: ESTs H40247 T25043"
     repeat_region   823. .880
                     /note="29 copies 2 mer aa 71% conserved"
     repeat_region   1468. .1513
                     /note="23 copies 2 mer aa 74% conserved"
     repeat_region   2074. .2177
                     /note="2 copies 52 mer 86% conserved"
     repeat_region   3183. .3362
                     /note="3 copies 60 mer 96% conserved"
     repeat_region   3533. .3766
                     /note="MIR repeat: matches 7. .237 of consensus"
     repeat_region   3846. .3966
                     /note="MIR repeat: matches 3. .139 of consensus"
     repeat_region   4059. .4194
                     /note="AluJb repeat: matches 1. .142 of consensus"
     repeat_region   4195. .4506
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   4507. .4684
                     /note="AluJb repeat: matches 142. .299 of consensus"
     repeat_region   6554. .6726
                     /note="AluJo repeat: matches 86. .262 of consensus"
     misc_feature    complement(7505. .7727)
                     /gene="dJ782L23.1"
                     /note="match: GSS: Em:AQ064456"
                     /evidence=not_experimental
     repeat_region   7728. .9025
                     /note="L1M4 repeat: matches 14. .1352 of consensus"
     repeat_region   9026. .9279
                     /note="L1M4 repeat: matches 1. .302 of consensus"
     repeat_region   9281. .9320
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   9321. .9497
                     /note="20 copies 2 mer at 85% conserved"
     repeat_region   9532. .9611
                     /note="AluJb repeat: matches 125. .300 of consensus"
     repeat_region   9646. .10125
                     /note="L1M4 repeat: matches 1466. .1553 of consensus"
     repeat_region   10164. .10429
                     /note="L1PA14 repeat: matches 5665. .6148 of consensus"
     repeat_region   10441. .10920
                     /note="L1M4 repeat: matches 987. .1206 of consensus"
     repeat_region   10921. .11230
                     /note="L1MB7 repeat: matches 1603. .2053 of consensus"
     repeat_region   11231. .14323
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   14324. .14604
                     /note="L1MB7 repeat: matches 2053. .4281 of consensus"
                     /note="L1MC2 repeat: matches 6052. .6330 of consensus"
```

```
                                        2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
                                          :::||::||::||||||::::|||||||||||
                                  98627 AAGCTGGGCGTTCCAGGGTTGCCTTACGATAT 98659

                   seq_name: gb_htg4:AC011241

                   seq_documentation_block:
                   LOCUS       AC011241   187493 bp    DNA                      HTG       29-OCT-1999
                   DEFINITION  Homo sapiens chromosome 2 clone NH0467P13 map unknown, WORKING
                               DRAFT SEQUENCE, in unordered pieces.
                   ACCESSION   AC011241
                   VERSION     AC011241.2  GI:6139206
                   KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
                   SOURCE      human.
                   ORGANISM    Homo sapiens
                               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                               Eutheria; Primates; Catarrhini; Hominidae; Homo.
                   REFERENCE   1  (bases 1 to 187493)
                   AUTHORS     Sulston,J.E. and Waterston,R.
                   TITLE       Toward a complete human genome sequence
                   JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
                   MEDLINE     99063792
                   REFERENCE   2  (bases 1 to 187493)
                   AUTHORS     Waterston,R.
                   TITLE       The sequence of Homo sapiens unknown clone NH0467P13
                   JOURNAL     Unpublished
                   REFERENCE   3  (bases 1 to 187493)
                   AUTHORS     Waterston,R.H.
                   TITLE       Direct Submission
                   JOURNAL     Submitted (04-OCT-1999) Genome Sequencing Center, Washington
                               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                               MO 63108, USA
                   COMMENT     On Oct 29, 1999 this sequence version replaced gi:6007883.
                               SUBMITTED BY: WUGSC
                               Genome Sequencing Center
                               Department of Genetics
                               Washington University
                               St. Louis MO 63108, USA
                               http://genome.wustl.edu/gsc
                               mailto:sapiens@watson.wustl.edu

                               NOTICE: This 'working draft' quality sequence may consist of
                               several contigs from automated sequence assembly concatenated
                               together.  No attempt has been made to order or orient the contigs
                               relative to one another correctly before concatenating.  At each
                               location in the sequence where contigs have been joined, several
                               consecutive Ns may have been inserted.

                               The attached annotation was produced using a purely automated
                               procedure.
                               * NOTE: This is a 'working draft' sequence.
                               * This record will be updated with the finished sequence
                               * as soon as it is available and the accession number will
                               * be preserved.
                   FEATURES             Location/Qualifiers
                   source          1..187493
                                   /organism="Homo sapiens"
                                   /db_xref="taxon:9606"
                                   /chromosome="2"
                                   /map="unknown"
                                   /clone="NH0467P13"
                                   /clone_lib="unknown"
                   repeat_region   323..621
                                   /rpt_family="Alu"
                   repeat_region   624..932
                                   /rpt_family="Alu"
                   exon            1427..1560
                   repeat_region   3133..3435
                                   /rpt_family="L2"
                   repeat_region   3452..3847
                                   /rpt_family="Retroviral"
                   repeat_region   3870..3961
                                   /rpt_family="(TA)n"
```

```
repeat_region   14605..15797
                /note="LIM7 repeat: matches 4281..5403 of consensus"
misc_feature    complement(15903..16103)
                /gene="dJ782L23.1"
                /note="match: STS: Em:AL009539"
                /evidence=not_experimental
repeat_region   18308..18645
                /note="LIM4 repeat: matches 2693..3042 of consensus"
repeat_region   18774..19928
                /note="LIM4 repeat: matches 3035..4224 of consensus"
repeat_region   19941..20261
                /note="MLT1A1 repeat: matches 16..362 of consensus"
repeat_region   20337..20897
                /note="LIM4 repeat: matches 4276..4883 of consensus"
repeat_region   20898..21146
                /note="LIMC4 repeat: matches 7729..7972 of consensus"
repeat_region   21147..21163
                /note="LIM4 repeat: matches 4883..4897 of consensus"
misc_feature    complement(21164..21779)
                /gene="dJ782L23.1"
                /note="match: GSS: Em:B76178.1"
                /evidence=not_experimental
misc_feature    21832..22423
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ029820.1"
                /evidence=not_experimental
misc_feature    21866..22366
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ532778.1"
                /evidence=not_experimental
misc_feature    complement(<22974..23194)
                /gene="dJ782L23.1"
                /note="match: 5' EST W24868 clone 308425"
repeat_region   23232..23389
                /note="MERSA repeat: matches 14..158 of consensus"
repeat_region   24137..24358
                /note="MIR repeat: matches 6..241 of consensus"
repeat_region   24511..24618
                /note="MIR repeat: matches 119..229 of consensus"
repeat_region   25107..27565
                /note="LIPA7 repeat: matches 3679..6145 of consensus"
repeat_region   29303..29618
                /note="MERB1B repeat: matches 2..337 of consensus"
misc_feature    30111..30371
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ174421"
                /evidence=not_experimental
misc_feature    join(30336..30371,30559..30682)
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ263653"
                /evidence=not_experimental
repeat_region   30372..30558
                /note="LIM4 repeat: matches 3080..3299 of consensus"
misc_feature    31605..32080
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ458966.1"
                /evidence=not_experimental
misc_feature    complement(33040..33387)
                /gene="dJ782L23.1"
                /note="match: GSS: Em:AQ196444"
                /evidence=not_experimental
repeat_region   33388..34107
```

```
alignment_scores:
   Quality:       37.00          Length:        11
   Ratio:          3.364         Gaps:           0
Percent Similarity: 100.000   Percent Identity: 63.636

alignment_block:
US-08-653-294-19 x HS782L23

Align seg 1/1  to: HS782L23  from: 1  to: 123925
```

```
repeat_region   4183..4223
                /rpt_family="(G)n"
repeat_region   4685..4829
                /rpt_family="L2"
repeat_region   4939..5132
                /rpt_family="MIR"
repeat_region   6310..6617
                /rpt_family="Alu"
repeat_region   6819..6847
                /rpt_family="(TG)n"
repeat_region   6902..7217
                /rpt_family="L1"
repeat_region   8297..8558
                /rpt_family="Alu"
repeat_region   8559..8579
                /rpt_family="(TAAA)n"
repeat_region   9186..9267
                /rpt_family="MIR"
repeat_region   10200..10358
                /rpt_family="MER1_type"
repeat_region   11030..11491
                /rpt_family="MaLR"
repeat_region   12118..12145
                /rpt_family="AT_rich"
repeat_region   12716..12900
                /rpt_family="(TTCC)n"
repeat_region   12948..12992
                /rpt_family="MER1_type"
repeat_region   12993..13181
                /rpt_family="MER53"
repeat_region   13182..13319
                /rpt_family="MER1_type"
repeat_region   13523..13706
                /rpt_family="MIR"
repeat_region   14375..14625
                /rpt_family="MaLR"
repeat_region   14751..14777
                /rpt_family="(CAAAAA)n"
repeat_region   15606..15793
                /rpt_family="MIR"
repeat_region   15977..16004
                /rpt_family="AT_rich"
repeat_region   16005..16208
                /rpt_family="MaLR"
repeat_region   16417..16475
                /rpt_family="L2"
repeat_region   16649..16910
                /rpt_family="MaLR"
repeat_region   17137..17442
                /rpt_family="MIR"
repeat_region   17543..17577
                /rpt_family="AT_rich"
repeat_region   17850..17875
                /rpt_family="AT_rich"
repeat_region   18650..18781
                /rpt_family="L2"
repeat_region   18812..19129
                /rpt_family="Alu"
repeat_region   19130..20452
                /rpt_family="L1"
repeat_region   20453..20639
                /rpt_family="MaLR"
repeat_region   20640..20939
                /rpt_family="Alu"
repeat_region   20940..21115
                /rpt_family="MaLR"
repeat_region   21478..21676
                /rpt_family="MER1_type"
repeat_region   21803..21873
                /rpt_family="L1"
repeat_region   22054..22338
                /rpt_family="Alu"
repeat_region   22378..24498
                /rpt_family="L2"

repeat_region   24499..24677
                /rpt_family="L1"
repeat_region   24678..25816
                /rpt_family="Alu"
repeat_region   25817..25913
                /rpt_family="L1"
repeat_region   25914..26034
                /rpt_family="Alu"
repeat_region   26035..26327
                /rpt_family="L1"
repeat_region   26328..28492
                /rpt_family="Alu"
repeat_region   28497..28786
                /rpt_family="L1"
repeat_region   28990..29344
                /rpt_family="Alu"
repeat_region   29776..29948
                /rpt_family="MER4-group"
repeat_region   29955..31168
                /rpt_family="MIR"
repeat_region   31330..31686
                /rpt_family="Retroviral"
repeat_region   31781..31870
                /rpt_family="MER1_type"
repeat_region   33879..33922
                /rpt_family="MIR"
repeat_region   34486..34595
                /rpt_family="(CGGG)n"
repeat_region   34617..34820
                /rpt_family="L2"
repeat_region   36144..36261
                /rpt_family="L2"
repeat_region   36432..36470
                /rpt_family="(TTCA)n"
repeat_region   37559..37609
                /rpt_family="L2"
repeat_region   37756..37786
                /rpt_family="(TAAAAA)n"
repeat_region   38373..38632
                /rpt_family="MER1_type"
repeat_region   39155..39253
                /rpt_family="MIR"
repeat_region   39281..39738
                /rpt_family="MaLR"
repeat_region   39743..39854
                /rpt_family="MIR"
repeat_region   40716..40799
                /rpt_family="MIR"
repeat_region   41810..41969
                /rpt_family="MER1_type?"
repeat_region   42150..42189
                /rpt_family="MIR"
repeat_region   42191..42445
                /rpt_family="MaLR"
repeat_region   42449..42626
                /rpt_family="Alu"
repeat_region   42648..42952
                /rpt_family="Alu"
repeat_region   43728..43789
                /rpt_family="L2"


alignment_scores:
             Quality:  37.00                 Length:     12
               Ratio:  3.700                   Gaps:      0
  Percent Similarity: 83.333    Percent Identity: 75.000

alignment_block:
US-08-653-294-19 x AC011241

Align seg 1/1  to: AC011241  from: 1  to: 187493

    1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
```

Wed Feb 9 08:47:52 2000

us-08-653-294-19.rge

```
     120136 TACTTATTGCTATCACAAGAACAGCACTCAGATAT 120171

seq_name: gb_ba2:AF026541

seq_documentation_block:
LOCUS       AF026541     1692 bp    DNA            BCT       30-OCT-1998
DEFINITION  Mycobacterium tuberculosis CeoB (ceoB) gene, complete cds; and CeoC
            (ceoC) gene, partial cds.
ACCESSION   AF026541
VERSION     AF026541.1  GI:2582553
KEYWORDS
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1 (bases 1 to 1692)
  AUTHORS   Chen,P. and Bishai,W.R.
  TITLE     Novel selection for isoniazid (INH) resistance genes supports a
            role for NAD+-binding proteins in mycobacterial INH resistance
  JOURNAL   Infect. Immun. 66 (11), 5099-5106 (1998)
  MEDLINE   99003115
REFERENCE   2 (bases 1 to 1692)
  AUTHORS   Chen,P. and Bishai,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1997) Molecular Microbiology and Immunology,
            Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205,
            USA
FEATURES             Location/Qualifiers
     source          1..1692
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     gene            681..1364
                     /gene="ceoB"
     CDS             681..1364
                     /gene="ceoB"
                     /note="TrkA homolog; complements the INH-sensitive
                     phenotype of oxyR deletion mutant of E. coli"
                     /codon_start=1
                     /transl_table=11
                     /product="CeoB"
                     /protein_id="AAC69359.1"
                     /db_xref="GI:2582554"
                     /translation="MRVVVMGCGRVGASVADGLSRIGHEVAIIDRDSAAFNRLSPQFA
                     GERVLGQGFDRDVLLRAGIQGADAFAAVSSGDNNIISARLARETFGVPRVVARIYDA
                     KRAEVYERLGIPITTVPWTTDRLLNALMQDTETAKWRDPTGTVVAEVVLHEDWVGH
                     RATDLEQATGARIAFLIRFGTGVLPEPKTVLQAGDKVYIAAISGRAAEAAAIAALPPS
                     EDFESGARR"
     gene            1361..>1692
                     /gene="ceoC"
     CDS             1361..>1692
                     /gene="ceoC"
                     /note="TrkA homolog; complements the INH-sensitive
                     phenotype of oxyR deletion mutant of E. coli"
                     /codon_start=1
                     /transl_table=11
                     /product="CeoC"
                     /protein_id="AAC69360.1"
                     /db_xref="GI:2582555"
                     /translation="MKVAVAGAGAVGRSVTRELVENGHDITLIERNPDHLDAAAIPEA
                     HWRLGDACELSLLESIHLEEFDVVVAATGDDKVNVVLSLLAKTEFAVPRVVARVNDPR
                     NEWLFNDA"
BASE COUNT      337 a    544 c    550 g     261 t
ORIGIN

alignment_scores:
        Quality:   36.00              Length:      12
          Ratio:    4.000               Gaps:       0
          Percent Similarity: 75.000    Percent Identity: 66.667
```

```
alignment_block:
  US-08-653-294-19 x AF026541

Align seg 1/1 to: AF026541   from:  1  to: 1692

     1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
       ||||||  |||||| |||::||:||||||||
  1252 TACAGCCGGCGATAAGGTCTATATCGCTGCGATAT 1287
```

OM of: US-08-653-294-19 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000 1:28 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgnl_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPEXT=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-08-653-294-19
Query length: 12
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 12596042
Search time (sec): 590.520000

score_list:
Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:T26331	+	36.00	129.80	12.50	382	! Human gene signature HUMGS08571
N_Geneseq_36:T76573	+	36.00	116.81	66.13	1728	! Pig alpha 1-6 fucosyltransfera
N_Geneseq_36:V11310	+	35.00	142.32	2.54	58	! D-A42d RNA sequence. L-nucleic
N_Geneseq_36:V74678	+	35.00	112.44	115.91	1847	! Staphylococcus aureus contig s
N_Geneseq_36:T66241	+	35.00	111.84	125.23	1981	! Romaine lettuce violaxanthin
N_Geneseq_36:Q44280	+	34.00	113.16	105.69	1092	! Pseudomonas cepacia DSM 3401 l
N_Geneseq_36:X12978	+	34.00	108.13	201.46	1959	! Enterococcus faecalis genome c
N_Geneseq_36:X13540	+	34.00	106.14	260.08	2469	! Enterococcus faecalis genome c
N_Geneseq_36:V19375	+	33.00	100.65	525.78	3003	! Hereditary multiple exostose a
N_Geneseq_36:X13357	+	33.00	97.94	744.22	4114	! Enterococcus faecalis genome c
N_Geneseq_36:X13127	+	33.00	89.16	2.3e+03	11410	! Enterococcus faecalis genome
N_Geneseq_36:T00127	+	32.00	115.67	76.58	337	! Hepatitis GB virus (HGBV) clone
N_Geneseq_36:T00049	+	32.00	115.67	76.58	337	! Hepatitis GB virus (HGBV) clone
N_Geneseq_36:T41700	+	32.00	103.75	353.42	1347	! Brugia pahangi beta-tubulin cD
N_Geneseq_36:Q12225	+	32.00	103.59	360.38	1371	! Octopus rhodopsin membrane pro
N_Geneseq_36:V21209_11	+	32.00	65.86	4.4e+04	110000	! Continuation (12 of 17) of
N_Geneseq_36:V80367	-	31.00	126.72	18.56	60	! Heteropolymer oligo used in immu
N_Geneseq_36:V80372	-	31.00	126.30	19.59	63	! Phosphorylated oligo used in a h
N_Geneseq_36:T27693	-	31.00	118.01	56.70	165	! pGEM32f target fragment. Cleava
N_Geneseq_36:T27694	-	31.00	116.10	72.44	206	! pGEM32f target fragment. Cleava
N_Geneseq_36:T70340	+	31.00	116.10	72.44	206	! DNA substrate for 5' nuclease.
N_Geneseq_36:T76640	+	31.00	116.10	72.44	206	! Duplex sequence used as 5' nucl
N_Geneseq_36:V53874	+	31.00	116.10	72.44	206	! Nucleotide sequence of the 5' n
N_Geneseq_36:V65800	+	31.00	116.10	72.44	206	! 5' Nuclease substrate DNA. Dete
N_Geneseq_36:V63426	+	31.00	116.10	72.44	206	! 206-mer duplex used as a substr
N_Geneseq_36:Q80775	+	31.00	115.23	81.02	228	! 5' Nuclease substrate. 5' Nucle
N_Geneseq_36:V16481	+	31.00	110.18	154.85	410	! DNA encoding a Bacillus thuring
N_Geneseq_36:V16485	+	31.00	110.18	154.85	410	! DNA encoding a Bacillus thuring
N_Geneseq_36:X21079	-	31.00	109.71	164.47	433	! Polynucleotide sequence from th
N_Geneseq_36:T28565	-	31.00	103.80	351.23	861	! Bacterial antibiotic resitance
N_Geneseq_36:Q24873	+	31.00	101.07	498.30	1182	! Alpha-ald gene. Accelerated be
N_Geneseq_36:X20675	-	31.00	100.75	518.81	1226	! Polynucleotide sequence from t
N_Geneseq_36:N82205	+	31.00	100.50	536.12	1263	! Nad B gene encoding L aspartat
N_Geneseq_36:N60877	+	31.00	99.61	600.66	1400	! Sequence encoding Serratia pho
N_Geneseq_36:T70130	+	31.00	99.39	618.21	1437	! Max-interacting protein coding
N_Geneseq_36:Q33064	+	31.00	96.58	885.54	1990	! Encodes Babesia bovis 60kD imm
N_Geneseq_36:T18995	+	31.00	96.58	885.54	1990	! Babesia merozoite surface prot
N_Geneseq_36:X39653	+	31.00	93.75	1.3e+03	2765	! Renal cancer associated gene.
N_Geneseq_36:Q12366	-	31.00	93.06	1.4e+03	2996	! Gene encoding enzyme with star
N_Geneseq_36:V31822	+	31.00	92.23	1.5e+03	3300	! Mutant Aspergillus oryzae DEBY
N_Geneseq_36:T59579	+	31.00	92.19	1.6e+03	3315	! Penicillium chrysogenum acetam
N_Geneseq_36:V16517	+	31.00	91.84	1.6e+03	3453	! DNA encoding a Bacillus thu
N_Geneseq_36:T61556	+	31.00	91.31	1.7e+03	3672	! Human signal mediator prote
N_Geneseq_36:V38466	+	31.00	90.87	1.8e+03	3867	! DNA encoding a Bacillus thu
N_Geneseq_36:X20635	-	31.00	89.67	2.1e+03	4444	! Polynucleotide sequence fro

seq_name: N_Geneseq_36:T26331

seq_documentation_block:
ID   T26331 standard; cDNA to mRNA; 382 BP.
AC   T26331;
DT   16-OCT-1996 (first entry)
DE   Human gene signature HUMGS08571
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 2058-2059; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   382 BP;   118 A;   72 C;   85 G;   102 T;

alignment_scores:
                      Quality: 36.00   Length: 12
                      Ratio: 4.000   Gaps: 0
Percent Similarity: 75.000   Percent Identity: 58.333

alignment_block:
US-08-653-294-19 x T26331

Align seg 1/1 to: T26331   from: 1 to: 382

      1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         |||||| |:: |||||:||||||||||
     36 TACAGGCGTGAACTACCGGCCTGGCCCTCAGATAT 71

seq_name: N_Geneseq_36:T76573

seq_documentation_block:
ID   T76573 standard; cDNA to mRNA; 1728 BP.
AC   T76573;
DT   05-MAR-1998 (first entry)
DE   Pig alpha 1-6 fucosyltransferase gene.
KW   Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
KW   guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW   GlcNAc; cancer diagnosis; ss.
OS   Sus scrofa.
FH   Key   Location/Qualifiers

FT　CDS　　　　　1..1728
FT　　　　　　　　/tag= a
PN　WO9727303-A1.
PD　31-JUL-1997.
PF　23-JAN-1997; J00171.
PR　22-JUL-1996; JP-192260.
PR　24-JAN-1996; JP-010365.
PR　21-JUN-1996; JP-161648.
PR　24-JUN-1996; JP-162813.
PA　(TOYM) TOYO BOSEKI KK.
PI　Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
DR　WPI: 97-393690/36.
DR　P-PSDB; W22124.
PT　Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for
PT　synthesis and modification of sugar chains and used as an antigen
PT　for production of diagnostic antibodies
PS　Claim 5; Page 30-34; 61pp; Japanese.
CC　T76573 and T76574 represent the coding sequences for the pig and human
CC　alpha 1-6 fucosyltransferases of the invention, respectively. The encoded
CC　enzyme transfers fucose from guanosine diphosphate to the 6-hydroxyl
CC　group of the GlcNAc nearest to R in the receptor molecule: (GlcNAcbeta
CC　1-2Manalpha 1-6)(GlcNAcbeta 1-2Manalpha 1-3)Manbeta 1-4GlcNAcbeta
CC　1-4GlcNAc-R to give (GlcNAcbeta 1-2Manalpha 1-6)(GlcNAcbeta 1-2Manalpha
CC　1-3)Manbeta 1-4GlcNAcbeta 1-4(Fucalpha 1-6)GlcNAc-R. It has an optimum pH
CC　of about 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10
CC　after 5 hours at 4 degrees C. The optimum working temperature of the
CC　enzyme is 30-37 degrees C. A bivalent metal is not required for activity
CC　of the enzyme, and the enzyme is not inhibited in the presence of 5 mM
CC　EDTA. The enzyme is useful in the synthesis and modification of sugar
CC　chains, and as antigen for the production of antibodies recognising the
CC　enzyme. The antibodies can be used for the diagnosis of cancer and other
CC　diseases.
SQ　Sequence　1728 BP;　521 A;　362 C;　419 G;　426 T;

alignment_scores:
　　　Quality:　36.00　　　　　　Length:　11
　　　Ratio:　3.273　　　　　　　Gaps:　0
Percent Similarity: 100.000　　Percent Identity: 72.727

alignment_block:
US-08-653-294-19 x T76573　　..

Align seg 1/1　to: T76573　from: 1 to: 1728

　2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
　　|||||:::|||||:||||:::|||||||||||
　723 CGCCTTGGAATCTCACAATTGGCGCTACGCTAC 755

seq_name: N_Geneseq_36:V11310

seq_documentation_block:
ID　V11310 standard: RNA; 58 BP.
AC　V11310;
DT　31-JUL-1998 (first entry)
DE　D-A42d RNA sequence.
KW　L-nucleic acid: target: D-nucleic acid: optical antipode; therapy;
KW　diagnosis; biosensor; herbicide; food additive; analysis; perfume;
KW　flavouring; cosmetic: purification: dialysis; separation; tumour;
KW　viral infection: bacterial infection; ss.
OS　Unknown.
FH　Key　Location/Qualifiers
FT　misc_binding　　.8
FT　　　　　　　　　/tag= a
FT　　　　　　　　　/note= "Binds to nucleotides 51. .58"
FT　misc_binding　　9. .10
FT　　　　　　　　　/tag= b
FT　　　　　　　　　/note= "Binds to nucleotides 46. .47"
FT　stem_loop　　　16. .26
FT　　　　　　　　　/tag= c
FT　stem_loop　　　29. .45
FT　　　　　　　　　/tag= d
FT　misc_binding　　46. .47

FT　　　　　　　　　/tag= e
FT　　　　　　　　　/note= "Binds to nucleotides 9. .10"
FT　misc_binding　　51. .58
FT　　　　　　　　　/tag= f
FT　　　　　　　　　/note= "Binds to nucleotides 1. .8"
PN　WO9808856-A2.
PD　05-MAR-1998.
PF　29-AUG-1997; E04726.
PR　30-AUG-1996; EP-113953.
PA　(BALD/) BALD R.
PA　(ERDM/) ERDMANN V A.
PA　(FUER/) FUERSTE J P.
PI　Bald R, Erdmann VA, Fuerste JP;
DR　WPI: 98-179376/16.
PT　L-nucleic acids that bind target molecules selectively - isolated by
PT　screening D-nucleic acids against optical antipode, useful in
PT　therapy and diagnosis
PS　Disclosure: Fig 13: 89pp; German.
CC　This sequence is used in a method which results in the production of
CC　L-nucleic acids able to bind to a target molecule. The method involves
CC　generating a heterogeneous population of D-nucleic acids then treating
CC　them with optical antipodes. Such nucleic acids are are useful in therapy
CC　and diagnosis, as e.g. biosensors, herbicides, food additives, in the
CC　analysis of perfumes or flavourings and in cosmetics, for formulating
CC　e.g. sunscreens or anti-wrinkle creams. They may stimulate or inhibit
CC　the function of the target molecule and can be coupled to markers or
CC　cytotoxins, carriers for affinity purification of the target molecule,
CC　including separation of enantiomers or elimination of enantiomeric
CC　impurities, or they can be used for purification of cellular factors or
CC　cells, e.g. the separation of toxic components by dialysis. The nucleic
CC　acids can be used in therapeutic applications for e.g. for tumours,
CC　viral and bacterial infection and for hypertension. The method allows
CC　selection and evaluation of high affinity nucleic acids. D-nucleic acids
CC　and optical antipodes of the target are more stable under physiological
CC　conditions than compounds with the natural configuration (particularly
CC　in the case of RNA). This stability means that potentially harmful
CC　metabolites are not formed and immune responses are not induced.
SQ　Sequence　58 BP;　13 A;　14 C;　13 G;　18 U;

alignment_scores:
　　　Quality:　35.00　　　　　　Length:　12
　　　Ratio:　3.889　　　　　　　Gaps:　0
Percent Similarity: 75.000　　Percent Identity: 50.000

alignment_block:
US-08-653-294-19 x V11310　　..

Align seg 1/1　to: V11310　from: 1 to: 58

　1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
　　　|||||:::::::　　||||||　:::|||||||
　6 UACCCCAAAGCGUUUUCGCCAUACCUAUUCGUUAU 41

seq_name: N_Geneseq_36:V74678

seq_documentation_block:
ID　V74678 standard: DNA: 1847 BP.
AC　V74678;
DT　16-MAR-1999 (first entry)
DE　Staphylococcus aureus contig SEQ ID #367.
KW　Computer readable medium: vaccine: S.aureus infection: immunodetection;
KW　cellulitis: eyelid infection; food poisoning: osteomyelitis; therapy;
KW　skin infection: surgical wound infection; scalded skin syndrome;
KW　toxic shock syndrome: ds.
OS　Staphylococcus aureus.
FH　Key　　　Location/Qualifiers
FT　misc_feature　　1381. .1440
FT　　　　　　　　　/tag= a
FT　　　　　　　　　/note= "these bases represent a line of missing text in
FT　　　　　　　　　the sequence listing in the specification. They
FT　　　　　　　　　are included to maintain the nucleotide numbering
FT　　　　　　　　　given in the specification for this DNA sequence"

PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1: Page 1260-1261; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 1847 BP;   627 A;   257 C;   334 G;   567 T;

alignment_scores:
     Quality: 35.00   Length: 10
       Ratio: 3.500   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 80.000

alignment_block:
US-08-653-294-19 x V74678

Align seg 1/1 to: V74678 from: 1 to: 1847

      3 LeuAlaIle***ArgIleAlaLeuArgTyr 12
        ::|||||||||:::|||||||||||||
    101 GTTGCTATTTCTAAGATTGCATTGCGTTAT 130

seq_name: N_Geneseq_36:T66241

seq_documentation_block:
ID   T66241 standard; CDNA; 1981 BP.
AC   T66241;
DT   28-JUL-1997 (first entry)
DE   Romaine lettuce violaxanthin de-epoxidase cDNA.
KW   Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW   photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW   xanthophyll; lettuce; ss.
OS   Lactuca sativa L. cv. romaine.
FH   Key   Location/Qualifiers
FT   misc_difference 26..29
FT   /*tag= a
FT   /note= "bases 26-29 are illegible in Fig 1"
FT   misc_difference 66..72
FT   /*tag= b
FT   /note= "bases 66-72 are illegible in Fig 1"
FT   misc_difference 105..110
FT   /*tag= c
FT   /note= "bases 105-110 are illegible in Fig 1"
FT   misc_difference 147..149
FT   /*tag= d
FT   /note= "bases 147-149 are illegible in Fig 1"
FT   misc_difference 186..189
FT   /*tag= e

FT   misc_difference 226..227
FT   /*tag= f
FT   /note= "bases 226-227 are illegible in Fig 1"
FT   cds   235..1656
FT   /*tag= g
FT   transit_peptide 235..609
FT   /*tag= h
FT   mat_peptide   610..1653
FT   /*tag= 1
PN   WO9717447-A2.
PD   15-MAY-1997.
PF   07-NOV-1996; U18291.
PR   07-NOV-1995; US-006315.
PR   06-AUG-1996; US-023502.
PA   (CALJ) CALGENE INC.
PI   Bugos RC, Rockholm DC, Yamamoto HY;
DR   WPI; 97-281036/25.
DR   P-PSDB; W09874.
PN   W09874.
PT   DNA encoding plant violaxanthin de-epoxidase - used to modify the
PT   sensitivity of a plant to light
PS   Claim 3; Fig 1; 41pp; English.
CC   A cDNA clone (T66241) codes for the 55 kDa violaxanthin de-epoxidase
CC   (VDE) (W09874) of romaine lettuce. VDE was purified from romaine
CC   lettuce chloroplasts and 2 tryptic peptides were used to develop
CC   primers (see also T66244-45), which amplified a partial VDE
CC   sequence. The amplified sequence was then used to screen a lettuce
CC   cDNA library, and the 1981 bp DNA sequence was identified. VDE
CC   nucleic acids (see also T66242-43), in sense or antisense
CC   orientation, can be used in genetic constructs, pref. also contg. a
CC   plastid translocation sequence, to modify VDE levels in plants.
CC   Increased levels result in the plant being tolerant of increased
CC   light and therefore more productive and/or more resistant to
CC   disease. Underexpression of VDE increases photosynthetic
CC   efficiency under low light. The photosensitivity of a range of
CC   crops, trees and ornamentals can be modified.
SQ   Sequence 1981 BP;   608 A;   337 C;   433 G;   577 T;

alignment_scores:
     Quality: 35.00   Length: 10
       Ratio: 3.889   Gaps: 0
Percent Similarity: 90.000   Percent Identity: 90.000

alignment_block:
US-08-653-294-19 x T66241

Align seg 1/1 to: T66241 from: 1 to: 1981

      3 LeuAlaIle***ArgIleAlaLeuArgTyr 12
        |||||||||||||||||| ||||||||||
    434 CTAGCCATTGCAAGGATAAAATCTCAGATAT 463

seq_name: N_Geneseq_36:Q44280

seq_documentation_block:
ID   Q44280 standard; DNA; 1092 BP.
AC   Q44280;
DT   09-DEC-1993 (first entry)
DE   Pseudomonas cepacia DSM 3401 lipD gene.
KW   Lipase; LipD; lipase modulator; limD; chaperone molecule;
KW   lipolysis; detergent; ss.
OS   Pseudomonas cepacia.
FH   Key   Location/Qualifiers
FT   unsure   391
FT   /*tag= a
FT   unsure   480
FT   /*tag= b
FT   unsure   481
FT   /*tag= c
FT   unsure   712..714
FT   /*tag= d
FT   unsure   734

```
FT
PN  WO9313200-A.     /*tag= e
PD  08-JUL-1993.
PF  18-DEC-1992; DK0391.
PR  20-DEC-1991; WO-DK0402.
PA  (NOVO ) NOVO-NORDISK AS.
PI  Buckley CM, Didderichsen BK, Hobson A, Joergensen ST;
DR  WPI; 93-227318/28.
DR  P-PSDB; R39396.
PT  Prepn. of active lipase in high quantities - by subjecting to
PT  denaturation and restructuring in presence of chaperone molecule
PS  Example 10; Page 43; 78pp; English.
CC  Two genes were cloned and sequenced from Pseudomonas cepacia DSM
CC  3401. The genes were designated lipD (Q44280) and limD (Q44281) and
CC  they code for a lipase and a lipase modulator protein, respectively.
CC  Due to the extreme GC content of the DNA, the sequence was difficult
CC  to determine (hence the "Others" in the sequence). The limD start
CC  codon is positioned 3 bp downstream of the lipD stop codon. LipD and
CC  LimD were found to be homologous to LipA and LimA, respectively. In
CC  denaturation/renaturation experiments, LimA chaperone protein was
CC  able to produce active LipD.
SQ  Sequence  1092 BP;  188 A;  353 C;  378 G;  166 T;

alignment_scores:
        Quality:    34.00         Length:    10
        Ratio:       4.250        Gaps:       0
Percent Similarity: 80.000    Percent Identity: 70.000

alignment_block:
US-08-653-294-19 x Q44280/rev

Align seg 1/1  to reverse of: Q44280  from: 1  to: 1092

        3 LeuAlaIle***ArgIleAlaLeuArgTyr 12
          |||::|| ||| |||||||||||
      960 CTGGCCGTACAGCCGCGCTCACTTCGATAC 931

seq_name: N_Geneseq_36:X12978

seq_documentation_block:
ID  X12978 standard; DNA; 1959 BP.
AC  X12978;
DT  19-MAR-1999  (first entry)
DE  Enterococcus faecalis genome contig SEQ ID NO:41.
KW  Enterococcus faecalis; contig; detection; Enterococcal infection;
KW  vaccine; attenuation; computer readable medium; ds.
OS  Enterococcus faecalis.
PN  WO9850555-A2.
PD  12-NOV-1998.
PF  04-MAY-1998; U08985.
PR  14-NOV-1997; US-066009.
PR  06-MAY-1997; US-044031.
PR  16-MAY-1997; US-046655.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
DR  WPI; 99-045171/04.
PT  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
PS  Claim 1; Page 413-414; 2084pp; English.
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  X12938 to X13919 represent these nucleotide sequences which are primary
CC  nucleotide sequences, also known as contigs. The computer-based system
CC  can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
SQ  Sequence  1959 BP;  542 A;  341 C;  322 G;  749 T;

alignment_scores:
        Quality:    34.00         Length:    12
        Ratio:       3.778        Gaps:       0
Percent Similarity: 75.000    Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x X12978

Align seg 1/1  to: X12978  from: 1  to: 1959

        1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
          ||| ||||||||||| |||::|| |||||||||||
     1358 TATAACTTGCTATATTAGAGATTTTTCTTCGGTAT 1393

seq_name: N_Geneseq_36:X13540

seq_documentation_block:
ID  X13540 standard; DNA; 2469 BP.
AC  X13540;
DT  19-MAR-1999  (first entry)
DE  Enterococcus faecalis genome contig SEQ ID NO:603.
KW  Enterococcus faecalis; contig; detection; Enterococcal infection;
KW  vaccine; attenuation; computer readable medium; ds.
OS  Enterococcus faecalis.
PN  WO9850555-A2.
PD  12-NOV-1998.
PF  04-MAY-1998; U08985.
PR  14-NOV-1997; US-066009.
PR  06-MAY-1997; US-044031.
PR  16-MAY-1997; US-046655.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
DR  WPI; 99-045171/04.
PT  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
PS  Claim 1; Page 1853-1855; 2084pp; English.
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  X12938 to X13919 represent these nucleotide sequences which are primary
CC  nucleotide sequences, also known as contigs. The computer-based system
CC  can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
SQ  Sequence  2469 BP;  771 A;  471 C;  483 G;  738 T;

alignment_scores:
        Quality:    34.00         Length:    12
        Ratio:       3.091        Gaps:       0
Percent Similarity: 91.667    Percent Identity: 50.000

alignment_block:
US-08-653-294-19 x X13540/rev

Align seg 1/1  to reverse of: X13540  from: 1  to: 2469

        1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
```

```
1322  :::::|||  :::::||||||||:::::||||||||||
      TTCCAGCTTCTTCTAACGCGGGGTTCGGCTTCGGTAT 1287

seq_name: N_Geneseq_36:V19375

seq_documentation_block:
ID  V19375 standard: cDNA; 3003 BP.
AC  V19375;
DT  20-AUG-1998  (first entry)
DE  Hereditary multiple exostose associated EXT2 gene isoform encoding cDNA.
KW  Hereditary multiple exostose; EXT2; chondrosarcoma; human; isoform;
KW  treatment: ss.
OS  Homo sapiens.
FH  Key             Location/Qualifiers
FT  CDS             167..2353
FT                  /tag= a
FT                  /Product= "EXT2 gene isoform"
PN  EP-837127-A2.
PD  22-APR-1998.
PF  26-AUG-1997; 306503.
PR  21-OCT-1996; CN-121928.
PA  (UYHU-) UNIV HUNAN MEDICAL.
PI  Deng HX, Fan CH, Xia J;
DR  WPI: 98-219110/20.
DR  P-PSDB: W44851.
PT  Cloned human EXT2 gene - associated with hereditary multiple
PT  exostoses or chondrosarcoma
PS  Claim 4; Pages 28-31; 31pp; English.
CC  This cDNA encodes an isoform of the EXT2 gene associated with hereditary
CC  multiple exostoses and chondrosarcoma. The polynucleotide is an isoform
CC  of the EXT2 gene described in Nature Genet., 14, 25, 1996. The
CC  polynucleotide can be used in the detection and treatment of EXT2-related
CC  diseases, and to identify compounds which activate or inhibit receptors
CC  for the encoded polypeptide. The polypeptide can be recombinantly
CC  produced by transforming or transfecting a cell with a vector containing
CC  the encoding nucleic acid.
SQ  Sequence 3003 BP;   751 A;   730 C;   766 G;   756 T;

alignment_scores:
      Quality:  33.00          Length:     11
      Ratio:   3.000            Gaps:        0
Percent Similarity: 100.000   Percent Identity: 54.545

alignment_block:
US-08-653-294-19 x V19375   ..

Align seq 1/1  to: V19375  from: 1  to: 3003

      1 TyrArgLeuAlaIle***ArgIleAlaLeuArg 11
        |||::||||||::::||||||::::|||:::
   1170 TACAAGCTGGCTGTGTCCCGGTTGTCATTGCAG 1202

seq_name: N_Geneseq_36:X13357

seq_documentation_block:
ID  X13357 standard; DNA; 4114 BP.
AC  X13357;
DT  19-MAR-1999  (first entry)
DE  Enterococcus faecalis genome contig SEQ ID NO:420.
KW  Enterococcus faecalis; contig; detection; Enterococcal infection;
KW  vaccine; attenuation; computer readable medium; ds.
OS  Enterococcus faecalis.
PN  WO9850555-A2.
PD  12-NOV-1998.
PF  04-MAY-1998; U08985.
PR  14-NOV-1997; US-066009.
PR  06-MAY-1997; US-044031.
PR  16-MAY-1997; US-046655.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
DR  WPI: 99-045171/04.
PT  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
PS  Claim 1; Page 1639-1641; 2084pp; English.
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  X12938 to X13919 represent these nucleotide sequences which are primary
CC  nucleotide sequences, also known as contigs. The computer-based system
CC  can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
SQ  Sequence 4114 BP;  1334 A;   661 C;   889 G;  1226 T;

alignment_scores:
      Quality:  33.00          Length:     11
      Ratio:   3.667            Gaps:        0
Percent Similarity: 81.818   Percent Identity: 54.545

alignment_block:
US-08-653-294-19 x X13357   ..

Align seq 1/1  to: X13357  from: 1  to: 4114

      2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
        |||::||||||:::   ||||:::|||  ||||||
   1110 AGAATTGCCTTACGACGATTAGCTAGTCGCTAT 1142

seq_name: N_Geneseq_36:X13127

seq_documentation_block:
ID  X13127 standard; DNA; 11410 BP.
AC  X13127;
DT  19-MAR-1999  (first entry)
DE  Enterococcus faecalis genome contig SEQ ID NO:190.
KW  Enterococcus faecalis; contig; detection; Enterococcal infection;
KW  vaccine; attenuation; computer readable medium; ds.
OS  Enterococcus faecalis.
PN  WO9850555-A2.
PD  12-NOV-1998.
PF  04-MAY-1998; U08985.
PR  14-NOV-1997; US-066009.
PR  06-MAY-1997; US-044031.
PR  16-MAY-1997; US-046655.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
DR  WPI: 99-045171/04.
PT  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
PS  Claim 1; Page 1021-1027; 2084pp; English.
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  X12938 to X13919 represent these nucleotide sequences which are primary
CC  nucleotide sequences, also known as contigs. The computer-based system
CC  can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
```

Wed Feb 9 08:47:53 2000

Left column:

```
SQ   Sequence   11410 BP;   3374 A;   2328 C;   1678 G;   4014 T;

alignment_scores:
         Quality:    33.00          Length:    11
         Ratio:      3.300          Gaps:       0
Percent Similarity: 90.909   Percent Identity: 63.636

alignment_block:
US-08-653-294-19 x X13127/rev

Align seg 1/1  to reverse of: X13127  from: 1  to: 11410

       2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         :::  ||||||||||:::::::||||||||||
    2924 AAATGTGCCATTACGAGAGTTTGCCTACGATAC 2892

seq_name: N_Geneseq_36:T00049

seq_documentation_block:
ID   T00049 standard; DNA; 337 BP.
AC   T00049;
DT   02-JUL-1996 (first entry)
DE   Hepatitis GB virus (HGBV) clone 50.
KW   Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW   reagents; non-A; non-B; non-C; non-D; non-E; clone 50;
KW   tamarin; infected plasma; lambda phage; cDNA library; ss.
OS   Hepatitis GB virus.
FH   Key             Location/Qualifiers
FT   mat_peptide     1..335
FT                   /*tag= a
FT                   /label= R81447
FT   mat_peptide     2..336
FT                   /*tag= b
FT                   /label= R81448
FT   mat_peptide     3..337
FT                   /*tag= c
FT                   /label= R81449
FT   mat_peptide     complement (1..335)
FT                   /*tag= d
FT                   /label= R81450
FT   mat_peptide     complement (2..336)
FT                   /*tag= e
FT                   /label= R82064
FT   mat_peptide     complement (3..337)
FT                   /*tag= f
FT                   /label= R82065
PN   WO9521922-A2.
PD   17-AUG-1995.
PF   14-FEB-1995; U02118.
PR   14-FEB-1994; US-196030.
PR   13-MAY-1994; US-242654.
PR   29-JUL-1994; US-283314.
PR   23-NOV-1994; US-344190.
PR   27-JAN-1995; US-344557.
PA   (ABBO ) ABBOTT LAB.
PI   Buijk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
PI   Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI   Simons JN;
DR   WPI; 95-293123/38.
DR   P-PSDB; R81447, R81448, R81449, R81450, R81449, R82064, R82065.
PT   Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT   for diagnosis and therapy of hepatitis GB virus
PS   Example 5; Page 193; 661pp; English.
CC   Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC   infected tamarin plasma, using standard procedures, was used to
CC   prepare a lambda phage HGBV cDNA library. The cDNA clone T00049,
CC   which encodes the proteins R81447-50 and R82064/65 (the 6 possible
CC   reading frames), was rescued from the lambda phage, searched
CC   against a sequence database and found to be an unique HGBV
CC   sequence. Reagents which comprise the HGBV DNA, or its protein
CC   prods. can be used for the diagnosis, therapy or in a vaccine to
```

Right column:

```
CC   prevent HGBV infection.
SQ   Sequence   337 BP;   99 A;   88 C;   79 G;   71 T;

alignment_scores:
         Quality:    32.00          Length:    12
         Ratio:      3.200          Gaps:       0
Percent Similarity: 83.333   Percent Identity: 50.000

alignment_block:
US-08-653-294-19 x T00049

Align seg 1/1  to: T00049  from: 1  to: 337

       1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         |||  ::||||||||::::::||||||||||
     243 TATCTTCTGCGGATGTCACGAGTGGCAATAAAGTAC 278

seq_name: N_Geneseq_36:T00127

seq_documentation_block:
ID   T00127 standard; DNA; 337 BP.
AC   T00127;
DT   02-JUL-1996 (first entry)
DE   Hepatitis GB virus (HGBV) clone.
KW   Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW   reagents; non-A; non-B; non-C; non-D; non-E; clone;
KW   tamarin; infected plasma; lambda phage; cDNA library; ss.
OS   Hepatitis GB virus.
PN   WO9521922-A2.
PD   17-AUG-1995.
PF   14-FEB-1995; U02118.
PR   14-FEB-1994; US-196030.
PR   13-MAY-1994; US-242654.
PR   29-JUL-1994; US-283314.
PR   23-NOV-1994; US-344190.
PR   23-NOV-1994; US-344185.
PR   27-JAN-1995; US-344557.
PA   (ABBO ) ABBOTT LAB.
PI   Buijk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
PI   Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI   Simons JN;
DR   WPI; 95-293123/38.
PT   Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT   for diagnosis and therapy of hepatitis GB virus
PS   Example 9; Page 286; 661pp; English.
CC   Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC   infected tamarin plasma, using standard procedures, was used to
CC   prepare a lambda phage HGBV cDNA library. cDNA clones rescued from
CC   the lambda phage were searched against a sequence database and
CC   found to be unique HGBV sequences. Then a series of PCR
CC   experiments were performed to obtain additional HGBV sequences,
CC   i.e. T00126-28. Reagents which comprise the HGBV DNA, or its
CC   protein prods. can be used for the diagnosis, therapy or in a
CC   vaccine to prevent HGBV infection.
SQ   Sequence   337 BP;   99 A;   88 C;   79 G;   71 T;

alignment_scores:
         Quality:    32.00          Length:    12
         Ratio:      3.200          Gaps:       0
Percent Similarity: 83.333   Percent Identity: 50.000

alignment_block:
US-08-653-294-19 x T00127

Align seg 1/1  to: T00127  from: 1  to: 337

       1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
         |||  ::||||||||::::::||||||||||
     243 TATCTTCTGCGGATGTCACGAGTGGCAATAAAGTAC 278

seq_name: N_Geneseq_36:T41700
```

```
seq_documentation_block:
ID   T41700 standard; DNA; 1347 BP.
AC   T41700;
DT   19-JAN-1997  (first entry)
DE   Brugia pahangi beta-tubulin cDNA.
KW   Filariasis; nematode; parasite; beta-tubulin; immunogen; vaccine;
KW   ss.
OS   Brugia pahangi.
PN   WO9632132-A1.
PD   17-OCT-1996.
PF   10-APR-1996; U04838.
PR   10-APR-1995; US-420982.
PA   (UYMC-) UNIV MCGILL.
PA   (UPJO ) UPJOHN CO.
PI   Bughio N, Feubert GM, Geary T, Prichard RK;
DR   WPI; 96-476844/47.
DR   P-PSDB; R99425.
PT   New vaccines for filarial parasite infection(s) - comprising
PT   C-terminal beta-tubulin amino acid sequence from a parasite
PS   Example 6; Page 51-53; 67pp; English.
CC   A cDNA clone (T41700) codes for the beta-tubulin (R99425) of
CC   the filarial nematode Brugia pahangi. The C-terminal portion
CC   (see also R99420) of the beta-tubulin is useful in novel vaccines
CC   against filarial parasite infections.
SQ   Sequence 1347 BP;    381 A;    262 C;    337 G;    367 T;

alignment_scores:
          Quality:     32.00          Length:      10
           Ratio:      4.000           Gaps:        0
Percent Similarity:   80.000    Percent Identity:  70.000

alignment_block:
US-08-653-294-19 x T41700        ..

Align seg 1/1  to: T41700   from: 1  to: 1347

     1 TyrArgLeuAlaIle***ArgIleAlaLeu 10
       |||||||||  |||  |||||||||::|||
   998 TACAGATTAAGAATTCATCGTATTCGTTG 1027


seq_name: N_Geneseq_36:Q12225


seq_documentation_block:
ID   Q12225 standard; DNA; 1371 BP.
AC   Q12225;
DT   13-SEP-1991  (first entry)
DE   Octopus rhodopsin membrane protein.
KW   Octopus; rhodopsin; membrane; helix; OR; ss.
FH   Key         Location/Qualifiers
FT   cds         1..1368
FT               /*tag=. a
FT               /product= membrane protein
FT   misc_feature 322..327
FT               /*tag= a
FT               /label= SphI
FT   misc_feature 592..597
FT               /note="restriction enzyme site"
FT               /*tag= b
FT               /label= MluI
FT   misc_feature 884..889
FT               /note="restriction enzyme site"
FT               /*tag= c
FT               /label= NdeI
FT   misc_feature 1069..1074
FT               /note="restriction enzyme site"
FT               /*tag= d
FT               /label= NheI
FT   misc_feature 1371
FT               /note="restriction enzyme site"
FT               /*tag= e
FT               /label= BamHI
FT               /note= "restriction enzyme site"
FT   PN   J03123486-A.
PN   27-MAY-1991.
PD   06-OCT-1989; 260261.
PF   06-OCT-1989; JP-260261.
PA   (HITA ) HITACHI KK.
DR   WPI; 91-197925/27.
DR   P-PSDB; R13362.
PT   Genetic engineering of membrane protein - by division of protein
PT   into cartridge genes corresp. to helix structure polypeptide(s)
PS   Disclosure; Fig 1; 17pp; Japanese.
CC   A gene cassette is prepd. by division of the OR membrane protein
CC   encoding helix structure polypeptides. Restriction sites are indicated.
CC   A base sequence contg. such a sequence downstream to the tryptophan
CC   regulating gene derived from the E. coli tryptophan operon, the trpL,
CC   or the trpE polypeptide translation regulating base sequence,
CC   and the N-terminal Met of the trpL (or trpE) or trpE polypeptide,
CC   respectively, is introduced in an expression vector for
CC   transformation of host cells.
SQ   Sequence 1371 BP;    298 A;    368 C;    351 G;    354 T;

alignment_scores:
          Quality:     32.00          Length:      10
           Ratio:      3.556           Gaps:        0
Percent Similarity:   90.000    Percent Identity:  60.000

alignment_block:
US-08-653-294-19 x Q12225/rev     ..

Align seg 1/1  to reverse of: Q12225  from: 1  to: 1371

     2 ArgLeuAlaIle***ArgIleAlaLeuArg 11
       ||:::||::||  |||:::|::||||||||
   1272 CGGGTAGCCCCTGCGGCGGGTAGCCCCTGAGG 1243
```

OM of: US-08-653-294-19 TO: EST:*   out_format : pfs
Date: Feb 8, 2000 4:03 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160700_15770/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
-NCPU=3 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-19
Query length: 12
Database: EST:*
Database sequences: 4538634
Database length: 188783192
Search time (sec): 8553.360000

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_gss8:AQ058120	+	38.00	144.64	27.25	468	AQ058120 CIT-HSP-2343C10.TF CIT
gb_est44:AV395907	+	38.00	144.50	27.78	476	AV395907 Chlamydomonas
gb_est20:AA897834	-	38.00	142.83	34.40	575	AA897834 NCP5E1T7 Perithecial N
gb_gss3:B02055	-	37.00	145.58	24.17	263	B02055 cSRL-147D11-u cSRL flow
gb_est5:H72837	+	37.00	139.84	50.47	504	H72837 ys06c01.s1 Soares fetal
gb_gss5:AQ815582	-	37.00	137.98	64.04	622	AQ815582 HS_5348_A2_E06_SP6E RF
gb_gss6:AQ851612	+	37.00	136.64	76.05	724	AQ851612 CpG1352B CpIOWAgDNAl C
gb_gss6:AQ864396	+	37.00	135.83	84.42	794	AQ864396 nbeb0023D13f CUGI Rice
gb_est12:AA320981	-	36.00	141.03	43.30	275	AA320981 EST23406 Adipose tissu
gb_gss1:CNS00WT6	+	36.00	135.87	84.02	494	ALO94008 Arabidopsis thaliana g
gb_est24:AI235399	+	36.00	135.72	85.56	502	AI235399 EST231961 Normalized r
gb_gss10:AQ210986	-	36.00	135.72	85.56	502	AQ210986 HS_3229_A2_D06_MR CIT
gb_gss9:AQ129594	+	36.00	135.62	86.72	508	AQ129594 HS_2254_A2_B12_MR CIT
gb_est27:AI407553	-	36.00	135.57	87.30	511	AI407553 EST235843 Normalized r
gb_est44:AW179172	+	36.00	135.41	89.05	520	AW179172 MR4-ST0069-080999-003-
gb_est10:AA194167	+	36.00	133.83	109.06	622	AA194167 zr37h06.r1 Soares_NhHM
gb_est35:AI828628	-	36.00	133.85	140.70	779	AI828628 wc10f08.x1 NCI_CGAP_Pr
gb_est35:AI864155	+	35.00	138.12	62.91	239	AI864155 wg64d07.x1 Soares_NSF_
gb_est41:AW159602	+	35.00	137.11	71.62	268	AW159602 zb04f06.x50 Xenopus ES
gb_gss11:AQ279730	+	35.00	135.88	83.83	308	AQ279730 CITBI-E1-2513B14.TR CI
gb_est1:T04318	-	35.00	133.58	112.69	400	T04318 365 Lambda-PRL2 Arabidop
gb_est8:AA003505	+	35.00	132.88	123.26	433	AA003505 T3197 MVAT4 bloodstrea
gb_est4:R65508	-	35.00	132.70	126.17	442	R65508 14012 Lambda-PRL2 Arabid
gb_est21:AA951497	-	35.00	132.68	126.49	443	AA951497 LD31979.5prime LD Dros
gb_est38:AL120179	+	35.00	132.31	132.65	462	AL120179 DKFZp761C237_r1 761 (s
gb_gss13:AQ438867	-	35.00	131.75	142.44	492	AQ438867 HS_5077_B2_H02_T7A RPC
gb_gss13:AA802027	+	35.00	131.59	145.39	501	AA802027 GM03744.5prime GM Dros
gb_gss13:AQ429279	-	35.00	131.38	149.34	513	AQ429279 CITBI-E1-2565K5.TR CIT
gb_gss4:AQ677348	+	35.00	131.21	152.64	523	AQ677348 HS_5526_A1_C10_T7A RPC
gb_gss10:AQ225567	+	35.00	131.21	152.64	523	AQ225567 HS_2012_A2_D03_MR CIT
gb_est14:AA439960	+	35.00	130.94	157.93	539	AA439960 LD14681.5prime LD Dros
gb_est32:AI729220	-	35.00	130.73	162.25	552	AI729220 BNLGHi12904 Six-day Co
gb_gss15:AQ591579	-	35.00	130.37	169.92	575	AQ591579 HS_5320_A2_B12_T7A RPC
gb_gss14:AQ569045	-	35.00	130.06	176.96	596	AQ569045 HS_5344_A1_F08_T7A RPC
gb_est40:AW142967	+	35.00	129.87	181.34	609	AW142967 EST293259 Normalized r
gb_est26:AI726854	+	35.00	129.72	184.71	619	AI726854 BNLGHi16722 Six-day Cot
gb_gss15:AQ662341	+	35.00	129.36	193.51	645	AQ662341 HS_5448_A2_B02_SP6E RF
gb_gss6:AQ840156	+	35.00	128.88	205.78	681	AQ840156 nbxb0052C09f CUGI Rice
gb_est25:AI258829	+	35.00	128.30	221.58	727	AI258829 ui21f07.y1 Soares mous
gb_gss11:AQ327491	-	35.00	127.49	245.88	797	AQ327491 nbxb0041M03r CUGI Rice
gb_gss4:AQ729455	+	35.00	127.39	249.02	806	AQ729455 HS_5474_A1_G10_T7AHS_5
gb_est26:AU005267	+	35.00	127.35	250.42	810	AU005267 Bombyx mori p

gb_gss5:AQ787789   -   35.00   126.48   280.01   894 ! AQ787789 HS_3072_A2_B11_T7C
gb_gss1:CNS0OMR8   +   34.00   133.61   112.23   249 ! AL080978 Arabidopsis thalian
gb_est21:AT000172   +   34.00   133.16   118.88   262 ! AT000172 Apple youn

seq_name: gb_gss8:AQ058120

seq_documentation_block:
LOCUS   AQ058120   468 bp   DNA   GSS   30-JUL-1998
DEFINITION CIT-HSP-2343C10.TF CIT-HSP Homo sapiens genomic clone 2343C10,
            genomic survey sequence.
ACCESSION   AQ058120
VERSION   AQ058120.1   GI:3354646
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 468)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
TITLE   Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CIT-HSP-2343C10.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2343C10"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT   144 a   101 c   82 g   141 t
ORIGIN

alignment_scores:
                Quality: 38.00   Length: 12
                Ratio: 4.222   Gaps: 0
      Percent Similarity: 75.000   Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x AQ058120

Align seg 1/1 to: AQ058120   from: 1   to: 468

      1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
        ||||||||||::::||| ||||| |||||
      1 TATCGAGCTCACTATAGGGCGAATTCGAGCTCCGTAC 36

seq_name: gb_est44:AV395907

seq_documentation_block:
LOCUS   AV395907   476 bp   mRNA   EST   09-DEC-1999
DEFINITION AV395907 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            cDNA clone CL51a05_r, mRNA sequence.
ACCESSION   AV395907
VERSION   AV395907.1   GI:6550123

```
KEYWORDS     EST.
SOURCE       Chlamydomonas reinhardtii
ORGANISM     Chlamydomonas reinhardtii
             Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
             Chlamydomonadaceae; Chlamydomonas.
REFERENCE    1 (bases 1 to 476)
AUTHORS      Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE        A Large Scale Structural Analysis of cDNAs in a Unicellular Green
             Alga, Chlamydomonas reinhardtii. I. Generation of 3451
             non-redundant Expressed Sequence Tags
JOURNAL      DNA Res. (1999) In press
COMMENT      On Jun 5, 1998 this sequence version replaced gi:3189716.
             Contact: Yasukazu Nakamura
             The First Laboratory for Plant Gene Research
             Kazusa DNA Research Institute
             Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
             Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES     Location/Qualifiers
     source  1..476
             /organism="Chlamydomonas reinhardtii"
             /strain="C9"
             /db_xref="taxon:3055"
             /clone="CL51a05_r"
             /clone_lib="Chlamydomonas reinhardtii C9"
             /dev_stage="photoautotrophic growth"
             /note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
             XhoI"
BASE COUNT   107 a   124 c   150 g   95 t
ORIGIN

alignment_scores:
             Quality:      38.00          Length:      10
             Ratio:        4.222          Gaps:         0
Percent Similarity: 90.000     Percent Identity: 80.000

alignment_block:
US-08-653-294-19 x AV395907
Align seg 1/1   to: AV395907   from: 1 to: 476

     3 LeuAlaIle***ArgIleAlaLeuArgTyr 12
       |||::::         ||||||||||||||
    72 CTGGCCTTGGTACGTATCCGCTCCAGGTAT 101

seq_name: gb_est20:AA897834

seq_documentation_block:
LOCUS        AA897834     575 bp      mRNA          EST      12-APR-1998
DEFINITION   NCP5EIT7 Perithecial Neurospora crassa cDNA clone NP5E1 3' end,
             mRNA sequence.
ACCESSION    AA897834
VERSION      AA897834.1  GI:3044267
KEYWORDS     EST.
SOURCE       Neurospora crassa.
ORGANISM     Neurospora crassa
             Eukaryota; Fungi; Ascomycota; Pyrenomycetes; Sordariales;
             Sordariaceae; Neurospora.
REFERENCE    1 (bases 1 to 575)
AUTHORS      Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
             Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
             Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K.,
             Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
             Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
             and Natvig,D.O.
TITLE        Expressed sequences from conidial, mycelial, and sexual stages of
             Neurospora crassa
JOURNAL      Fungal Genet. Biol. 21, 348-363 (1997)
MEDLINE      97435549
COMMENT      On Jan 19, 1998 this sequence version replaced gi:2151810.
             Contact: Natvig,D.O./Nelson,M.A.
             Department of Biology
             University of New Mexico
             Casteter Hall, Albuquerque, NM 87131, USA
             Tel: 505 277 3411
             Fax: 505 277 0304
             Email: ngp@biology.unm.edu
             Deposited in GSDB at the National Center for Genome Resources with
             accession GSDB:S:1146747
             Seq primer: T7.
FEATURES     Location/Qualifiers
     source  1..575
             /organism="Neurospora crassa"
             /strain="fl a"
             /db_xref="taxon:5141"
             /clone="NP5EI"
             /clone_lib="Perithecial"
             /sex="Mating type a (fluffy), fertilized"
             /tissue_type="Perithecia (fruiting bodies)"
             /dev_stage="perithecia"
             /note="mRNA isolated from 5 day old perithecia (fruiting
             bodies) of the fluffy strain fl a (Mating type a),
             fertilized with conidia from 74-OR23-IV A (Mating type A).
             cDNA directionally cloned into pBluescript SK(-) using
             the Uni-ZAP XR vector system (Stratagene, La Jolla, CA)."
BASE COUNT   189 a   115 c    90 g   181 t
ORIGIN

alignment_scores:
             Quality:      38.00          Length:      12
             Ratio:        3.800          Gaps:         0
Percent Similarity: 83.333     Percent Identity: 66.667

alignment_block:
US-08-653-294-19 x AA897834
Align seg 1/1   to: AA897834   from: 1 to: 575

     1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
       |||||||||        |||||||        |||||||
   102 TACAGACTGGATATATCAAAGTCAAGTTAAGGTAT 137

seq_name: gb_gss3:B02055

seq_documentation_block:
LOCUS        B02055       263 bp     DNA          GSS       13-JUL-1996
DEFINITION   cSRL-147D11-u cSRL flow sorted Chromosome 11 specific cosmid Homo
             sapiens genomic clone cSRL-147D11, genomic survey sequence.
ACCESSION    B02055
VERSION      B02055.1  GI:1411333
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 263)
AUTHORS      Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
             Jones,D., Ward,T., Gillilan,E., Schageman,J., Probst,S.,
             Harris,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M.,
             Kupfer,K. and Garner,H.R.
TITLE        Genomic Sequence Sampled Map of Chromosome 11
JOURNAL      Unpublished (1996)
COMMENT      Contact: Evans GA, Shane Probst
             McDermott Center for Human Growth and Development
             University of Texas Southwestern Medical Center At Dallas
             5323 Harry Hines Blvd, Dallas TX 75235-8591
             Tel: 214-648-1600
             Fax: 214-648-1666
             Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
             PCR PRimers
             FORWARD: TCTTGAATTTAGGAGTTGTC
             BACKWARD: GTTATGTCACCAGCTATTCC
             Seq primer: T7
             Class: cosmid ends
             High quality sequence stop: 263.
```

```
FEATURES
     source          Location/Qualifiers
                     1..263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CSRL-147D11"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
                     cosmid"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                     library prepared from flow sorted human Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT       64 a      43 c      55 g      87 t     14 others
ORIGIN

alignment_scores:
        Quality:     37.00                 Length:      12
        Ratio:       3.700                 Gaps:         0
Percent Similarity: 83.333      Percent Identity: 58.333

alignment_block:
US-08-653-294-19 x B02055/rev

Align seg 1/1  to reverse of: B02055  from: 1  to: 263

  1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
    ||| ||::::::   ||||::::||||||||
195 TATTCCCTAACTGTGTGAGGATTGGATTGAGAGATAT 160

seq_name: gb_est5:H72837

seq_documentation_block:
LOCUS       H72837             504 bp    mRNA        EST       27-OCT-1995
DEFINITION  ys06c01.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:213984 3', mRNA sequence.
ACCESSION   H72837
VERSION     H72837.1  GI:1044653
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97044478
COMMENT     On May 8, 1995 this sequence version replaced gi:799648.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 528
            High quality sequence stops: 380
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 528   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Homo sapiens"
```

```
                     /db_xref="GDB:3779288"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:213984"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI:
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAGAATAATTAAAGATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT73 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      120 a     103 c     125 g     147 t      9 others
ORIGIN

alignment_scores:
        Quality:     37.00                 Length:      12
        Ratio:       4.111                 Gaps:         0
Percent Similarity: 75.000      Percent Identity: 58.333

alignment_block:
US-08-653-294-19 x H72837/rev

Align seg 1/1  to reverse of: H72837  from: 1  to: 504

  1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
    ||| ||::::::   ||| :::: ||||||||
289 TACCCACTAGCCCTGAAAAGGGTAAAACTTCGATAT 254

seq_name: gb_gss5:AQ815582

seq_documentation_block:
LOCUS       AQ815582           622 bp    DNA                  GSS       26-AUG-1999
DEFINITION  HS_5348_A2_E06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-924 Col-12 Row-I, genomic survey sequence.
ACCESSION   AQ815582
VERSION     AQ815582.1  GI:5777975
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 622)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,R., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 924  row: I  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
```

```
source          1..622
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=924 Col=12 Row=I"
                /clone_lib="RPCI-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT       190 a     145 c     123 g     144 t      20 others
ORIGIN

alignment_scores:
        Quality:       37.00         Length:       12
          Ratio:       4.111           Gaps:        0
Percent Similarity:   75.000  Percent Identity:  66.667

alignment_block:
US-08-653-294-19 x AQ815582

Align seg 1/1  to: AQ815582  from: 1  to: 622

  1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
    |||||   |||||     :::||||||||||||
321 TATAGAACGGCCATAGAATAAGTGGCTCTAAGATAC 356

seq_name: gb_gss6:AQ851612

seq_documentation_block:
LOCUS       AQ851612      724 bp      DNA          GSS      18-OCT-1999
DEFINITION  CpG1352B CpIOWAgDNAl Cryptosporidium parvum genomic similar to SKB1
            homologue (negative regulator of mitosis) (regulator of Shk1, a
            p21(Cdc42/Rac)-activated kinase (PAK)), genomic survey sequence.
ACCESSION   AQ851612
VERSION     AQ851612.1  GI:6063307
KEYWORDS    GSS.
SOURCE      Cryptosporidium parvum.
ORGANISM    Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE   1  (bases 1 to 724)
AUTHORS     Strong,W.B. and Nelson,R.G.
TITLE       Cryptosporidium parvum GSS Project
JOURNAL     Unpublished (1997)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3553959.
            Contact: Nelson, R. G.
            Depts. of Medicine & Pharmaceutical Chemistry
            San Francisco General Hospital-University of California, San
            Francisco
            Box 0811, San Francisco, CA 94143-0811, USA
            Tel: 415 206 8846
            Fax: 415 206 3353
            Email: malaria@itsa.ucsf.edu
            For Annotation Data see http://medsfgh.ucsf.edu/id/CpTags/home.html
            Seq primer: T3
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..724
                     /organism="Cryptosporidium parvum"
                     /strain="IOWA"
                     /db_xref="taxon:5807"
                     /clone_lib="CpIOWAgDNAl"
                     /lab_host="E. coli XL2 Blue MRF'"
                     /note="Vector: pBlueScript II (SK-).; Site_1: EcoRV; C.
                     parvum (IOWA isolate) genomic DNA was hydrodynamically
                     sheared to produce fragments having a tight size
                     distribution between 2-4 kb by Dr. Yvonne Thorstenson of
                     the Stanford DNA Sequencing and Technology Center
                     (http://sequence.
                     www.stanford.edu/group/techdev/shear.htm). The randomly
                     sheared gDNA was chromatographed on Sephacryl S-400 to
                     remove any small fragments and DNA eluting in the void
                     volume was subcloned into an EcoR V-digested, alkaline
                     phosphatase-treated pBlueScript II (SK-) vector and
```

```
source          transformed into E. coli strain XL2 Blue MRF'.
                Recombinant clones from the first plating of the library
                were selected for sequence analysis using T3 and T7
                primers."

BASE COUNT       264 a      93 c     121 g     244 t       2 others
ORIGIN

alignment_scores:
        Quality:       37.00         Length:       11
          Ratio:       4.111           Gaps:        0
Percent Similarity:   81.818  Percent Identity:  81.818

alignment_block:
US-08-653-294-19 x AQ851612/rev

Align seg 1/1  to reverse of: AQ851612  from: 1  to: 724

  1 TyrArgLeuAlaIle***ArgIleAlaLeuArg 11
    ||||||||   |||  ||||||||||||
552 TATAGGCTCTAAATAAGACGTATAGCGTTGAGG 520

seq_name: gb_gss6:AQ864396

seq_documentation_block:
LOCUS       AQ864396      794 bp      DNA          GSS      03-NOV-1999
DEFINITION  nbeb0023D13f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0023D13f, genomic survey sequence.
ACCESSION   AQ864396
VERSION     AQ864396.1  GI:6214957
KEYWORDS    GSS.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 794)
AUTHORS     Wing,R.A. and Dean,R.A.
TITLE       A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence start: 84
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1..794
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0023D13f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
```

Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximatley 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

```
BASE COUNT      245 a    173 c    183 g    183 t    10 others
ORIGIN
```

```
alignment_scores:
       Quality:     37.00              Length:     11
         Ratio:      3.700               Gaps:      0
Percent Similarity:  90.909    Percent Identity:  81.818
```

```
alignment_block:
US-08-653-294-19 x AQ864396    ..
```

Align seq 1/1 to: AQ864396   from: 1 to: 794

```
         2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
           ||||:::||||:||||||||||||||||| |||||
        24 CGTCTCACTATAGCCGGAATTGCAGCCCGATAT 56
```

seq_name: gb_est12:AA320981

```
seq_documentation_block:
LOCUS       AA320981          275 bp    mRNA            EST        19-APR-1997
DEFINITION  EST23406 Adipose tissue, white II Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA320981
VERSION     AA320981.1  GI:1973309
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 275)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     12140200
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397790.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
```

```
FEATURES         Seq primer: M13 Reverse.
     source          Location/Qualifiers
                     1..275
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):121563"
                     /db_xref="taxon:9606"
                     /clone_lib="Adipose tissue, white II"
                     /tissue_type="adipose tissue, white"
                     /dev_stage="adult"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
```

```
BASE COUNT      80 a     48 c     61 g     82 t     4 others
ORIGIN
```

```
alignment_scores:
       Quality:     36.00              Length:     12
         Ratio:      3.273               Gaps:      0
Percent Similarity:  91.667    Percent Identity:  50.000
```

```
alignment_block:
US-08-653-294-19 x AA320981/rev    ..
```

Align seq 1/1 to reverse of: AA320981 from: 1 to: 275

```
         1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
           :::||||||:::::::     |||::::|::||||||:::
        38 TTCAGACTGTCAGTTTTTAGAATAGTGCTAAGGTTT 3
```

seq_name: gb_gss1:CNS00WT6

```
seq_documentation_block:
LOCUS       CNS00WT6                                     GSS       28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence SP6 end of BAC T12G12
            of TAMU library from strain Columbia of Arabidopsis thaliana,
            genomic survey sequence.
ACCESSION   AL094008
VERSION     AL094008.1  GI:5295162
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 494)
AUTHORS     Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 494)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr)
FEATURES         Location/Qualifiers
     source          1..494
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="TAMU"
                     /clone="T12G12"
                     /note="end : SP6"
```

```
DNA             494 bp    DNA
```

```
BASE COUNT      119 a    177 c    114 g     84 t
ORIGIN
```

```
alignment_scores:
       Quality:     36.00              Length:     11
         Ratio:      3.600               Gaps:      0
Percent Similarity:  90.909    Percent Identity:  45.455
```

alignment_block:

```
US-08-653-294-19 x CNS00WT6

Align seg 1/1  to: CNS00WT6  from: 1  to: 494

          2 ArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
            |||::|||::  :::::::|||::|||||
        416 CGAATGGCAGTAGTCAAGGTAGCCATGAGGTAC 448

seq_name: gb_est24:AI235399

seq_documentation_block:
LOCUS       AI235399       502 bp     mRNA             EST        31-JAN-1999
DEFINITION  EST231961 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVCR36 3' end, mRNA sequence.
ACCESSION   AI235399
VERSION     AI235399.1  GI:3828905
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 502)
AUTHORS     Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE       Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL     Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151609.
            Other_ESTs: TC63429
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..502
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2040962"
                     /db_xref="taxon:10118"
                     /clone="ROVCR36"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT    169 a    90 c    85 g   158 t
ORIGIN

alignment_scores:
     Quality:        36.00                     Length:       12
     Ratio:           4.000                     Gaps:          0
Percent Similarity:  75.000     Percent Identity: 75.000

alignment_block:
US-08-653-294-19 x AI235399/rev  .

Align seg 1/1   to reverse of: AI235399  from: 1  to: 502

          1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
            |||||||||||  |||  ||||
        499 TACAGACTTGCCTACAGCCTATCTTATTGAGGTAT 464

seq_name: gb_gss10:AQ210986

seq_documentation_block:
LOCUS       AQ210986       502 bp    DNA                 GSS        18-SEP-1998
DEFINITION  HS_3229_A2_D06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3229 Col=12 Row=G, genomic survey
            sequence.
ACCESSION   AQ210986
VERSION     AQ210986.1  GI:3619955
```

```
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3229 row: G  column: 12
            Class: BAC ends
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1..502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3229 Col=12 Row=G"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT    136 a    120 c    112 g    131 t     3 others
ORIGIN

alignment_scores:
     Quality:        36.00                     Length:       12
     Ratio:           4.500                     Gaps:          0
Percent Similarity:  66.667     Percent Identity: 58.333

alignment_block:
US-08-653-294-19 x AQ210986/rev  .

Align seg 1/1   to reverse of: AQ210986  from: 1  to: 502

          1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
            |||||||||  :::  |||||   |||||
        108 TATAGATTACAGGTAGAGAGAATTCGAGCTAGGTAC 73

seq_name: gb_gss9:AQ129594

seq_documentation_block:
LOCUS       AQ129594       508 bp    DNA                 GSS        23-SEP-1998
DEFINITION  HS_2254_A2_B12_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2254 Col=24 Row=C, genomic survey
            sequence.
ACCESSION   AQ129594
VERSION     AQ129594.1  GI:3506760
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 508)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
```

BASE COUNT   151 a   108 c   108 g   144 t
ORIGIN   Site_2: NotI"

alignment_scores:
      Quality: 36.00   Length: 12
      Ratio: 4.000   Gaps: 0
Percent Similarity: 75.000   Percent Identity: 75.000

alignment_block:
US-08-653-294-19 x AI407553/rev   ..

Align seg 1/1 to reverse of: AI407553 from: 1 to: 511

   1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
     |||||||||||| ||| ||| ||||||| ||||||||
 378 TACAGACTGGCCTACAGCCTTATCTTATTGAGGTAT 343

seq_name: gb_est44:AW179172

seq_documentation_block:
LOCUS   AW179172   520 bp   mRNA   EST   16-NOV-1999
DEFINITION MR4-ST0069-080999-003-C02 ST0069 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW179172
VERSION   AW179172.1 GI:6445209
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 520)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES
  TITLE   The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT   On Jul 7, 1999 this sequence version replaced gi:5406194.
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-ST0069-
            080999-003-C02&t3=1999-09-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 519.
FEATURES   Location/Qualifiers
     source   1..520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ST0069"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196.716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT   144 a   113 c   118 g   145 t
ORIGIN

alignment_scores:
      Quality: 36.00   Length: 12
      Ratio: 4.000   Gaps: 0
Percent Similarity: 75.000   Percent Identity: 50.000

alignment_block:

COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2254 row: C column: 24
            Class: BAC ends
            High quality sequence stop: 508.
FEATURES   Location/Qualifiers
     source   1..508
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2254 Col=24 Row=C"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT   175 a   99 c   91 g   141 t   2 others
ORIGIN

alignment_scores:
      Quality: 36.00   Length: 12
      Ratio: 3.273   Gaps: 0
Percent Similarity: 91.667   Percent Identity: 58.333

alignment_block:
US-08-653-294-19 x AQ129594/rev   ..

Align seg 1/1 to reverse of: AQ129594 from: 1 to: 508

   1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
     ||||||||:::::||||||::::|||::| |||
 378 TATAGGTTGTCAGTTACGAGGATGGCAATATTTAT 343

seq_name: gb_est27:AI407553

seq_documentation_block:
LOCUS   AI407553   511 bp   mRNA   EST   09-FEB-1999
DEFINITION EST235843 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVD217 3' end, mRNA sequence.
ACCESSION   AI407553
VERSION   AI407553.1 GI:4251057
KEYWORDS   EST.
SOURCE   Rattus sp.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 511)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE   Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   On Feb 17, 1998 this sequence version replaced gi:2889619.
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712; Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21
FEATURES   Location/Qualifiers
     source   1..511
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="ROVD217"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;

Wed Feb 9 08:47:59 2000

us-08-653-294-19.rst

US-08-653-294-19 x AW179172   . .

Align seg 1/1 to: AW179172 from: 1 to: 520

```
  1 TyrArgLeuAlaIle***ArgIleAlaLeuArgTyr 12
    ::::||| |||:::: |||::: ||||||||||
459 TTCAGATGCGCAGTTGATAGAGTACAGCTAAGATAT 494
```

OM protein - protein search, using sw model

```
            GenCore version 4.5
            Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:     February 8, 2000, 01:29:40 ; Search time 122.56 Seconds
                                         (without alignments)
                                         2.706 Million cell updates/sec

Title:          US-08-653-294-20
Perfect score:  63
Sequence:       1 YRLAIRIXRILLRY 14

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:    188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing:  Minimum Match 08
                  Listing first 45 summaries

Database :  A_Geneseq_36:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	42	66.7	12	1	R95429	HLA-B2702 84-79-84
2	42	66.7	12	1	W33798	Peptide B2702.84-7
3	42	66.7	12	1	W33799	Immunomodulating d
4	41	65.1	20	1	R92907	HLA-B2702 CTL modu
5	41	65.1	20	1	R95428	HLA-B2702 84-75-84
6	41	65.1	20	1	W33778	Immunomodulating d
7	37	58.7	469	1	W56793	L. lactis F1 porti
8	36	57.1	20	1	R92909	HLA-B2702 CTL modu
9	36	57.1	20	1	R92908	HLA-B2702 CTL modu
10	36	57.1	20	1	W33791	HLA-B2702 CTL modu
11	36	57.1	20	1	W33792	Peptide B2702.84-7
12	34	54.0	20	1	R95430	Peptide B2702.84-7
13	34	54.0	334	1	R95941	HLA-B2702 84-75T/7
14	34	54.0	334	1	W37094	Canine Y5 receptor
15	34	54.0	445	1	W15232	Canis domesticus Y
16	34	54.0	445	1	W15230	Rat neuropeptide Y
17	34	54.0	445	1	W27602	Rat neuropeptide Y
18	34	54.0	445	1	W27604	Rat neuropeptide Y
19	34	54.0	445	1	W27603	Human neuropeptide
20	34	54.0	455	1	R95939	Human Y5 receptor.
21	34	54.0	455	1	W29447	Human hippocampal
22	34	54.0	455	1	W29413	Human hippocampal
23	34	54.0	455	1	W37093	Homo sapiens hippo
24	34	54.0	456	1	R95940	Rat Y5 receptor. M
25	34	54.0	456	1	W29446	Rat hypothalamic n
26	34	54.0	456	1	W29412	Rat hypothalamic n
27	34	54.0	456	1	W37095	Canis domesticus Y
28	34	54.0	456	1	W37092	Rattus norvegicus
29	34	54.0	466	1	W15233	Mouse neuropeptide
30	34	54.0	893	1	W13899	Thermotoga neapoli
31	34	54.0	893	1	W13900	Thermotoga neapoli
32	34	54.0	893	1	W13901	Thermotoga neapoli
33	34	54.0	893	1	W13902	Thermotoga neapoli
34	34	54.0	893	1	W53918	Wild type Tne poly
35	34	54.0	893	1	W53919	Mutant Tne polymer
36	34	54.0	893	1	W78758	Tne DNA polymerase
37	34	54.0	893	1	W83974	Tne DNA polymerase
38	34	54.0	893	1	W83973	Tne DNA polymerase
39	34	54.0	893	1	W83980	Tne DNA polymerase
40	34	54.0	893	1	W83981	Tne DNA polymerase
41	34	54.0	893	1	W83982	Tne DNA polymerase
42	34	54.0	893	1	W83983	Tne DNA polymerase
43	34	54.0	893	1	W83984	Tne DNA polymerase
44	34	54.0	893	1	W83985	Tne DNA polymerase
45	34	54.0	893	1	W83976	Tne DNA polymerase

ALIGNMENTS

```
RESULT  1
R95429
ID   R95429 standard; peptide; 12 AA.
AC   R95429;
DT   12-NOV-1996  (first entry)
DE   HLA-B2702 84-79-84 palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; Cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C,  Krensky AM;
DR   WPI: 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-79-84 palindrome. These sequences can be used to isolate
CC   the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC   protein associated with T-cell activation in mammalian T-cells, and is
CC   also immunologically cross reactive with the heat shock protein Hsc70.
CC   p74 is found in a limited number of cell types, but is particularly
CC   expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC   cell with an amphoteric detergent, and then passed through an affinity
CC   column containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence  12 AA;

Query Match             66.7%;  Score 42;  DB 1;  Length 12;
Best Local Similarity   78.6%;  Pred. No. 0.045;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy          1 YRLAIRIXRILLRY 14
              ||||||  ||-|||
Db          1 YRLAIR--RIALRY 12

RESULT  2
W33798
ID   W33798 standard; peptide; 12 AA.
AC   W33798;
```

```
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84-79/79-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation: autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W3378-98 and W3378-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) or (aa84-79) (Laa77-76R); aa76 = E or V; aa77 =
CC   D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a hydrophobic or
CC   small amino acid; aa82 = R or L; aa83 = G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 12 AA;

Query Match             66.7%;  Score 42;  DB 1;  Length 12;
Best Local Similarity   78.6%;  Pred. No. 0.045;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  2;

Qy    1 YRLAIRIXRILLRY 14
         |||||| || |||
Db    1 YRLAIR--RIALRY 12


RESULT  3
W33799
ID   W33799 standard; peptide; 12 AA.
AC   W33799;
DT   19-JUN-1998 (first entry)
DE   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation: autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 17; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is
```

```
CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B =(R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 = E or
CC   V; aa77 = D, S or N; aa79 = R or G; aa80 = I or N; aa81, aa84= a
CC   hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. They can also inhibit the
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence 12 AA;

Query Match             66.7%;  Score 42;  DB 1;  Length 12;
Best Local Similarity   78.6%;  Pred. No. 0.045;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  2;

Qy    1 YRLAIRIXRILLRY 14
         |||||| || |||
Db    1 YRLAIR--RIALRY 12


RESULT  4
R92907
ID   R92907 standard; peptide; 20 AA.
AC   R92907;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence 20 AA;

Query Match             65.1%;  Score 41;  DB 1;  Length 20;
Best Local Similarity   55.0%;  Pred. No. 0.11;
Matches  11;  Conservative  1;  Mismatches  1;  Indels  2;  Gaps  6;  Gaps  1;

Qy    1 YRLAIRI-----XRILLRY 14
         ||||||:      ||||||
Db    1 YRLAIRLNERRENLRIALRY 20
```

```
RESULT   5
ID   R95428 standard; peptide; 20 AA.
AC   R95428;
DT   12-NOV-1996  (first entry)
DE   HLA-B2702 84-75-84 palindrome.
KW   HLA; p74; alpha-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.
OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example: Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of the
CC   human-leucocyte-associated antigens.  These sequences represents the
CC   HLA-B2702 84-75-84 palindrome.  These sequences can be used to isolate
CC   the protein p74 from a T-cell lysate. p74 is a T-cell surface membrane
CC   protein associated with T-cell activation in mammalian T-cells, and is
CC   also immunologically cross reactive with the heat shock protein Hsc70.
CC   p74 is found in a limited number of cell types, but is particularly
CC   expressed on B and T cells. p74 can be isolated by lysis of a suitable
CC   cell with an amphoteric detergent, and then passed through an affinity
CC   column containing a covalently bound HLA-B2702 palindromic peptide.
CC   Compositions comprising the extracellular fragment of p74 combined with
CC   HLA-B2702.60-84 (see R95416), induces calcium influx, and inhibits
CC   cytotoxic T lymphocyte (CTL) differentiation or cytolysis. Candidate
CC   compounds can be screened for their effect on the cytolytic activity of
CC   T-cells, by combining them with the extracellular portion of p74 and
CC   determining the amount of binding between the candidate compound and p74.
CC   Modulation of CTL activity can be inhibited in a cellular composition
CC   containing T-cells and antigen presenting cells (APCs), by adding to the
CC   mix the extracellular portion of p74, in an amount sufficient to compete
CC   with p74 for the binding of the p74 ligand.
SQ   Sequence  20 AA;

Query Match             65.1%;  Score 41; DB 1;  Length 20;
Best Local Similarity   55.0%;  Pred. No. 0.11;
Matches    11;  Conservative    1;  Mismatches    2;  Indels    6;  Gaps    1;

Qy      1 YRLAIRI------XRILLRY 14
            ||||||        || |||
Db      1 YRLAIRLNERRENLRIALRY 20

RESULT   6
ID   W33778 standard; peptide; 20 AA.
AC   W33778;
DT   19-JUN-1998  (first entry)
DE   Immunomodulating dimer peptide #1.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   Immunomodulating dimer; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Claim 16; Page 35; 41pp; English.
CC   This sequence represents a specifically claimed immunomodulating
CC   dimer peptide of the invention. A peptide-type compound or variant is
CC   claimed which has immunomodulating activity, including the N-terminal
CC   acylated and/or C-terminal amidated or esterified forms of up to 60
CC   amino acids, where the peptide-type compound comprises the formula; A-B,
CC   where A, B -(R a76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or
CC   V; aa77 - D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a
CC   hydrophobic or small amino acid; aa82 = R or L; aa83 = G or R; and aa
CC   represents amino acid. The sequence in the brackets may optionally be
CC   absent or truncated at any peptide type bond within the brackets. The
CC   compounds comprise amino acid sequences related to a Class I HLA-B
CC   alpha1 domain (positions 79-84). They can be used to inhibit cytotoxic
CC   T-lymphocytes (CTL) from undesirably attacking cells in a host or in
CC   vitro. They can also be used in combination with antigenic peptides or
CC   proteins of interest to activate CTLs. They can also inhibit the
CC   proliferation of T cells in response to anti-CD3. The peptide can be
CC   used for preventing rejection of transplants or for treating autoimmune
CC   diseases, e.g. diabetes, rheumatoid arthritis and lupus erythematosis.
CC   The products can also be used for detection and diagnosis.
SQ   Sequence  20 AA;

Query Match             65.1%;  Score 41; DB 1;  Length 20;
Best Local Similarity   55.0%;  Pred. No. 0.11;
Matches    11;  Conservative    1;  Mismatches    2;  Indels    6;  Gaps    1;

Qy      1 YRLAIRI------XRILLRY 14
            ||||||        || |||
Db      1 YRLAIRLNERRENLRIALRY 20

RESULT   7
ID   W56793 standard; Protein; 469 AA.
AC   W56793;
DT   04-AUG-1998  (first entry)
DE   L. lactis F1 portion of F0F1 ATPase beta subunit.
KW   Biomass production; uncoupled ATPase; F0F1 ATPase; membrane bound;
KW   F1; Lactococcus lactis.
OS   Lactococcus lactis.
PN   WO9810089-A1.
PD   12-MAR-1998.
PF   08-SEP-1997; DK0373.
PR   06-SEP-1996; DK-000963.
PA   (JENS/) JENSEN P R.
PI   Snoep JL, Westerhoff HV;
DR   WPI; 98-193637/17.
DR   N-PSDB; V29571.
PT   Method improving production of biomass or a desired product - by
PT   expressing an uncoupled ATPase activity in the cell
PS   Example 5; Pages 45-46; 78pp; English.
CC   This is the beta subunit of soluble part (F1) of membrane bound (F0F1
CC   type) H+-ATPase. The DNA encoding this or a portion of F1 exhibiting
CC   ATPase activity is derived from Lactococcus lactis subsp. cremoris strain
CC   MG1363 and is used in a novel method for improving the production of
CC   biomass or a desired product from a cell. The method comprises expressing
CC   an uncoupled ATPase activity in the cell to induce conversion of ATP to
CC   ADP without primary effects on other cellular metabolites or functions
CC   and incubating the cell with a suitable substrate to produce the biomass
CC   or product. The expression is directed using a vector including DNA
CC   encoding the soluble part (F1) of the membrane bound (F0F1 type)
CC   H+-ATPase or a portion of F1 exhibiting ATPase activity, the DNA being
CC   derived from Lactococcus lactis subsp. cremoris, Lactococcus lactis
CC   subsp. lactis, Streptococcus thermophilus, Phaffia rhodozyma or
CC   Trichoderma reesel, where the DNA is under the control of a promoter. An
CC   ideal ATPase is the membrane bound H+ ATPase. This enzyme complex
CC   consists of two parts, the membrane integral part and the (F0) and the
CC   cytoplasmic part (F1). Together the two parts couple the hydrolysis of
CC   ATP and ADP to the translocation of protons across the cytoplasmic
CC   membrane, or vice versa. The proton gradient is used to drive ATP
CC   synthesis from ADP and Pi. The method can be used for optimising the
```

```
CC   formation of biomass or a desired product, e.g. the product may be lactic
CC   acid which results in the acidification of dairy products. The method is
CC   more efficient than currently used methods of biomass production.
SQ   Sequence   469 AA;

Query Match             58.7%;  Score 37;  DB 1;  Length 469;
Best Local Similarity   42.9%;  Pred. No. 13;
Matches    6;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 YRLAIRIXRILLRY 14
              |:|::|:|:|| ||
Db        362 YEVAMEVQRVLQRY 375


RESULT   8
R92909
ID   R92909 standard; peptide; 20 AA.
AC   R92909;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84(T)).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO952697S-Al.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92913 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match             57.1%;  Score 36;  DB 1;  Length 20;
Best Local Similarity   50.0%;  Pred. No. 0.95;
Matches   10;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    6;

Qy          1 YRLAIRI-----XRILLRY 14
              ||||||:    || |||
Db          1 YRLAIRLNERRENLRTALRY 20


RESULT   9
R92908
ID   R92908 standard; peptide; 20 AA.
AC   R92908;
DT   16-MAY-1996  (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75(T)/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO952697S-Al.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
```

```
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts  - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens. This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702. These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant. This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments). The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match             57.1%;  Score 36;  DB 1;  Length 20;
Best Local Similarity   50.0%;  Pred. No. 0.95;
Matches   10;  Conservative    1;  Mismatches    1;  Indels    3;  Gaps    6;

Qy          1 YRLAIRI-----XRILLRY 14
              ||||||:    || |||
Db          1 YRLATRLNERRENLRIALRY 20


RESULT   10
W33791
ID   W33791 standard; peptide; 20 AA.
AC   W33791;
DT   19-JUN-1998  (first entry)
DE   Peptide B2702.84-75T/75-84 tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO974435l-Al.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-086530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula; A-B, where A, B =
CC   (R aa76-77L) (aa79-84) or (aa84-79) (Laa77-76R); aa76 - E or V; aa77 -
CC   D, S or N; aa79 - R or G; aa80 - I or N; aa81, aa84- a hydrophobic or
CC   small amino acid; aa82 - R or L; aa83 - G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alphal domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence   20 AA;
```

Query Match   57.1%;   Score 36; DB 1; Length 20;
Best Local Similarity 50.0%;   Pred. No. 0.95;
Matches 10; Conservative 1; Mismatches 1; Indels 3; Gaps 6;

Qy 1 YRLAIRI------XRILLRY 14
       |||| |: || |||
Db 1 YRLATRLNERRENLRIALRY 20

RESULT 11
W33792
ID   W33792 standard; peptide; 20 AA.
AC   W33792;
DT   19-JUN-1998 (first entry)
DE   Peptide B2702.84 75/75-84T tested for immunomodulating activity.
KW   Immunomodulating dimer; immunosuppressant drug; CTL activation;
KW   transplantation; autoimmune disease; Class I HLA-B alpha-1 domain;
KW   rejection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744351-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U08689.
PR   24-MAY-1996; US-653294.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Beulow R, Clayberger C, Krensky AM;
DR   WPI; 98-085530/08.
PT   New immunomodulating dimer peptide(s) - based on a Class I HLA-B
PT   alpha-1 domain, used for preventing rejection of transplants or
PT   treating autoimmune diseases
PS   Example 1; Page 19; 41pp; English.
CC   Peptides W33784-98 and W33778-9 were assayed for their immunomodulating
CC   activity. A peptide-type compound or variant is claimed which has
CC   immunomodulating activity, including the N-terminal acylated and/or
CC   C-terminal amidated or esterified forms of up to 60 amino acids, where
CC   the peptide-type compound comprises the formula: A-B, where A, B -
CC   (R aa76-77L) (aa79-84) or (Laa77-76R) (Laa77-76R); aa76 - E or V; aa77 -
CC   D, S or N: aa79 = R or G; aa80 - I or N; aa81, aa84= a hydrophobic or
CC   small amino acid: aa82 - R or L; aa83 - G or R; and aa represents amino
CC   acid. The sequence in the brackets may optionally be absent or truncated
CC   at any peptide type bond within the brackets. The compounds comprise
CC   amino acid sequences related to a Class I HLA-B alpha1 domain (positions
CC   79-84). They can be used to inhibit cytotoxic T-lymphocytes (CTL) from
CC   undesirably attacking cells in a host or in vitro. They can also be
CC   used in combination with antigenic peptides or proteins of interest to
CC   activate CTLs. They can also inhibit the proliferation of T cells in
CC   response to anti-CD3. The peptide can be used for preventing rejection
CC   of transplants or for treating autoimmune diseases, e.g. diabetes,
CC   rheumatoid arthritis and lupus erythematosis. The products can also be
CC   used for detection and diagnosis.
SQ   Sequence 20 AA;

Query Match   57.1%;   Score 36; DB 1; Length 20;
Best Local Similarity 50.0%;   Pred. No. 0.95;
Matches 10; Conservative 1; Mismatches 1; Indels 3; Gaps 6;

Qy 1 YRLAIRI------XRILLRY 14
       ||||||: | || |||
Db 1 YRLAIRLNERRENLRTALRY 20

RESULT 12
R95430
ID   R95430 standard; peptide; 20 AA.
AC   R95430;
DT   12-NOV-1996 (first entry)
DE   HLA-B2702 84-75T/75-84T palindrome.
KW   HLA; p74; alpha1-helix; human-leucocyte-associated antigen; inhibitor;
KW   T-cell lysate; membrane protein; mammal; heat shock protein; Hsc70; APC;
KW   B cell; calcium influx; cytotoxic T lymphocyte; CTL; differentiation;
KW   cytolysis; antigen presenting cell.

OS   Synthetic.
PN   WO9513288-A1.
PD   18-MAY-1995.
PF   10-NOV-1994; U12985.
PR   10-NOV-1993; US-150493.
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM;
DR   WPI; 95-194027/25.
PT   Compsns. comprising lymphoid surface membrane proteins - which may
PT   inhibit cytolytic activity and differentiation of CTLs.
PS   Example; Page 12; 29pp; English.
CC   R95413, and R95415-R95431 represent palindromes and fragments of
CC   human-leucocyte-associated antigens. This sequence represents the
CC   HLA-B2702 84-75T/75-84T palindrome. These sequences can be used to
CC   isolate the protein p74 from a T-cell lysate. p74 is a T-cell surface
CC   membrane protein associated with T-cell activation in mammalian T-cells,
CC   and is also immunologically cross reactive with the heat shock protein
CC   Hsc70. p74 is found in a limited number of cell types, but is
CC   particularly expressed on B and T cells. p74 can be isolated by lysis of
CC   a suitable cell with an amphoteric detergent, and then passed through an
CC   affinity column containing a covalently bound HLA-B2702 palindromic
CC   peptide. Compositions comprising the extracellular fragment of p74
CC   combined with HLA-B2702.60-84 (see R95416), induces calcium influx, and
CC   inhibits cytotoxic T lymphocyte (CTL) differentiation or cytolysis.
CC   Candidate compounds can be screened for their effect on the cytolytic
CC   activity of T-cells, by combining them with the extracellular portion of
CC   p74 and determining the amount of binding between the candidate compound
CC   and p74. Modulation of CTL activity can be inhibited in a cellular
CC   composition containing T-cells and antigen presenting cells (APCs), by
CC   adding to the mix the extracellular portion of p74, in an amount
CC   sufficient to compete with p74 for the binding of the p74 ligand.
SQ   Sequence 20 AA;

Query Match   54.0%;   Score 34; DB 1; Length 20;
Best Local Similarity 52.6%;   Pred. No. 2.2;
Matches 10; Conservative 1; Mismatches 2; Indels 6; Gaps 1;

Qy 1 YRLAIRI------XRILLR 13
       ||||||: | ||
Db 1 YRLAIRLNETRENLRIALR 19

RESULT 13
R95941
ID   R95941 standard; Protein; 334 AA.
AC   R95941;
DT   14-OCT-1996 (first entry)
DE   Canine Y5 receptor.
KW   Y5 receptor; atypical neuropeptide Y1 receptor; feeding behavior;
KW   G protein-coupled receptor; agonist; antagonist; obesity;
KW   bulimia; anorexia.
OS   Canis familiaris.
PN   WO9616542-A1.
PD   06-JUN-1996.
PF   01-DEC-1995; U15646.
PR   02-DEC-1994; US-349025.
PA   (SYNA-) SYNAPTIC PHARM CORP.
PI   Branchek T, Gerald CPG, Walker MW, Weinshank RL;
DR   WPI; 96-277371/28.
DR   N-PSDB; T30435.
PT   Modifying feeding behaviour using Y5 receptor (ant)agonists -
PT   increases or decreases food consumption, for treatment of e.g.
PT   obesity or bulimia
PS   Claim 58; Fig 15; 235pp; English.
CC   Canine Y5 receptor (R95941) was identified as the homologue of rat
CC   hypothalamic Y5 receptor (R95940), isolated as an 'atypical Y1
CC   receptor'. The receptor belongs to the G protein-coupled receptor
CC   superfamily. It is encoded by a cDNA clone (see also T30435) that
CC   was isolated by PCR amplification using primers (T30436-37) based
CC   on human and rat cDNA clones (T30433-34). Recombinant canine Y5
CC   receptor can be produced in prokaryotic or eukaryotic (e.g. COS,
CC   293 or Sf9 insect) host cells. It is used to identify Y5 ligands

CC   (agonists and antagonists) that can be used to treat obesity,
CC   bulimia or anorexia.
SQ   Sequence 334 AA;

Query Match   54.0%; Score 34; DB 1; Length 334;
Best Local Similarity 42.9%; Pred. No. 34;
Matches   6; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
         ||:|| | :||:|
Db   95 YRIAFTISLLLVQY 108

RESULT 14
W37094
ID   W37094 standard; Protein; 334 AA.
AC   W37094;
DT   08-JUN-1998 (first entry)
DE   Canis domesticus Y5 receptor.
KW   Y5 receptor; treatment; anorexia; bulimia; obesity;
KW   feeding behaviour; modification; atypical neuropeptide.
OS   Canis domesticus.
PN   WO9746250-A1.
PD   11-DEC-1997.
PF   04-JUN-1997; U09504.
PR   21-FEB-1997; US-803600.
PR   04-JUN-1996; US-666650.
PA   (SYNA-) SYNAPTIC PHARM CORP.
PI   Branchek T, Gerald CP, Walker MW, Weinshank RL;
DR   WPI: 98-051901/05.
DR   N-PSDB; V00639.
PT   DNA encoding canine hypothalamic atypical neuro:peptide Y/peptide YY
PT   receptor, Y5 - useful for identification of compounds which are
PT   capable of modifying feeding behaviour
PS   Claim 5; Fig 15; 273pp; English.
CC   The sequence is that of a Y5 receptor (Y5-R).
CC   Y5-R can be used in processes to determine whether a chemical compound
CC   specifically binds to and activates or inhibits a Y5-R by measuring a
CC   second messenger response. The chemical compounds can be used to
CC   increase or reduce the activity of a Y5-R. In particular, inhibitors
CC   can be used to treat obesity and activators can be used to treat
CC   anorexia. Antagonists capable of alleviating (by decreasing the
CC   activity of Y5-R) an abnormality can be identified by administering a
CC   potential antagonist to a transgenic mammal as above, and determining
CC   whether the substance alleviates the physical and behavioural
CC   abnormalities displayed by the transgenic mammal as a result of
CC   overactivity of a Y5-R. Agonists can be identified in a similar manner,
CC   but where the abnormality is alleviated by increasing the activity of
CC   Y5-R.
SQ   Sequence 334 AA;

Query Match   54.0%; Score 34; DB 1; Length 334;
Best Local Similarity 42.9%; Pred. No. 34;
Matches   6; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
         ||:|| | :||:|
Db   95 YRIAFTISLLLVQY 108

RESULT 15
W15232
ID   W15232 standard; Protein; 445 AA.
AC   W15232;
DT   21-JUL-1997 (first entry)
DE   Rat neuropeptide Y-Y5 receptor.
KW   Neuropeptide Y-Y5; appetite; obesity; G-protein coupled receptor;
KW   antiobesity; hypotensive; neuronal growth factor;
KW   cardiovascular drug; anti-psychotic; neuroleptic; antidiabetic;
KW   agonist; antagonist.
OS   Rattus sp.

PN   WO9717440-A1.
PD   15-MAY-1997.
PF   08-NOV-1996; AU0706.
PR   09-NOV-1995; AU-006467.
PA   (GARV-) GARVAN INST MEDICAL RES.
PI   Herzog H;
DR   WPI: 97-281029/25.
DR   N-PSDB; T66911.
DR   N-PSDB; T66910.
PT   DNA encoding the neuropeptide Y-Y5 receptor - for screening for
PT   NPY-Y5 antagonists and agonists, useful as anti-obesity agents,
PT   anti-hypertensive agents cardiovascular drugs, etc.
PS   Claim 17; Fig 3; 44pp; English.
CC   A novel rat neuropeptide Y (NPY)-Y1-like receptor (W15232),
CC   designated NPY-Y5 receptor, is a G-protein coupled receptor of
CC   NPY, which is involved in appetite/obesity regulation. Its amino
CC   acid sequence was deduced from a cDNA clone (T66911) isolated from
CC   a rat hypothalamic library. Human (W15230) and mouse (W15233)
CC   NPY-Y5 receptors have also been identified. The receptors can be
CC   expressed on the cell surface of host (pref. CHO, human embryonic
CC   kidney 293 or insect sf9) cells. The receptors or host cells can
CC   be used to screen for (ant)agonists of NPY useful as potential
CC   hypotensives, cardiovascular drugs, neuronal growth factors,
CC   anti-psychotic, anti-obesity or anti-diabetic drugs.
SQ   Sequence 445 AA;

Query Match   54.0%; Score 34; DB 1; Length 445;
Best Local Similarity 42.9%; Pred. No. 45;
Matches   6; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
         ||:|| | :||:|
Db   207 YRIAFTISLLLVQY 220

Search completed: February 8, 2000, 01:29:41
Job time: 1753 sec

OM protein - protein search, using sw model

```
               GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: February 7, 2000, 11:54:30 ; Search time 117.7 Seconds
                                          (without alignments)
                                          5.611 Million cell updates/sec

Title: US-08-653-294-20
Perfect score: 63
Sequence: 1 YRLAIRIXRILLRY 14

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters: 142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08
                 Listing first 45 summaries

Database :   PIR_62:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	39	61.9	485	2	T09974	H+-transporting AT
2	39	61.9	486	2	B70775	probable atpD prot
3	38	60.3	133	2	D72110	hypothetical prote
4	38	60.3	194	2	F64075	urease accessory p
5	38	60.3	478	2	H70474	ATP synthase F1 be
6	38	60.3	709	1	OXCKPM	acyl-CoA oxidase (
7	36	57.1	96	2	B69327	conserved hypothet
8	36	57.1	502	2	OXCKAX	acyl-CoA oxidase (
9	36	57.1	709	1	OXCKX4	acyl-CoA oxidase (
10	36	57.1	709	1	OXCKX	acyl-CoA oxidase (
11	36	57.1	3131	1	S39842	enniatin synthetas
12	35	55.6	724	2	OXCKP2	acyl-CoA oxidase (
13	35	55.6	724	2	JC4563	acyl-CoA oxidase (
14	34	54.0	157	2	F72612	hypothetical prote
15	34	54.0	276	2	F72721	probable citrate l
16	34	54.0	323	2	A70404	acetyl-CoA carboxy
17	34	54.0	662	1	OXCKX5	acyl-CoA oxidase (
18	34	54.0	2009	1	S49764	SEC7 protein - yea
19	33	52.4	149	2	G72250	ribosomal protein
20	33	52.4	260	2	S57939	AhyR protein - Aer
21	33	52.4	331	2	G75035	iron (iii) abc tra
22	33	52.4	470	2	S30597	H+-transporting AT
23	33	52.4	478	2	S37547	H+-transporting AT
24	33	52.4	485	2	A70706	probable phoR prot
25	33	52.4	532	2	G70986	probable coA ligas
26	33	52.4	1729	2	A49282	fusion protein 1a/
27	33	52.4	2376	2	S48405	probable membrane
28	33	52.4	2630	2	T08868	polyprotein P1 - A
29	32	50.8	205	2	T10296	fibroblast growth
30	32	50.8	339	2	D72509	hypothetical prote
31	32	50.8	489	2	T13026	hypothetical prote
32	32	50.8	591	2	B72086	hypothetical prote
33	32	50.8	785	1	T01541	hypothetical prote
34	31	49.2	113	1	R5BY1E	ribosomal protein
35	31	49.2	113	2	S55962	ribosomal protein
36	31	49.2	185	2	S74416	hypothetical prote
37	31	49.2	195	2	I50456	pituitary adenylat
38	31	49.2	220	2	G72472	hypothetical prote
39	31	49.2	224	2	F71329	probable phosphogl
40	31	49.2	277	2	JC2565	replication protei
41	31	49.2	294	2	H38888	COI intron 15 prot
42	31	49.2	331	2	A71128	probable iron (III
43	31	49.2	334	2	G69303	iron (III) ABC tra
44	31	49.2	343	2	S15949	hypothetical prote
45	31	49.2	351	2	H70570	hypothetical prote

ALIGNMENTS

RESULT   1
T09974
H+-transporting ATP synthase (EC 3.6.1.34) beta chain - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T09974
R;Robison, K.
submitted to the EMBL Data Library, September 1994
A;Reference number: Z16911
A;Accession: T09974
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-485 <ROB>
A;Cross-references: EMBL:U15186; NID:g699323; PID:g699347
C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match   61.9%;   Score 39;   DB 2;   Length 485;
Best Local Similarity   57.1%;   Pred. No. 3.7;
Matches   8; Conservative 2; Mismatches 2;   4; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
          ||:| : ||| ||
Db 376 YRVAQEVIRILQRY 389

RESULT   2
B70775
probable atpD protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Jul-1999
C;Accession: B70775
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 39, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:9829987
A;Accession: B70775
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-486 <COL>
A;Cross-references: GB:273419; GB:AL123456; NID:g3261573; PIDN:CAA97743.1; PID:e24365
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: atpD
C;Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase
F;197-366/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match   61.9%;   Score 39;   DB 2;   Length 486;
Best Local Similarity   57.1%;   Pred. No. 3.7;

Matches   8; Conservative   2; Mismatches   2; Indels   4; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           ||:| : ||| ||
Db   377 YRVAQEVIRILQRY 390

RESULT   3
D72110
hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C;Species: Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999
C;Accession: D72110
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606
A;Accession: D72110
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-133 <ARN>
A;Cross-references: GB:AE001604; GB:AE001363; NID:g4376438; PID:g4376450
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: CPn0181

Query Match   60.3%;   Score 38; DB 2; Length 133;
Best Local Similarity   61.5%;   Pred. No. 1.7;
Matches   8; Conservative   2; Mismatches   2; Indels   3; Gaps   0;

Qy   2 RLAIRIXRILLRY 14
           |: |:||||||
Db   88 RIPWRLKRILLRY 100

RESULT   4
F64075
urease accessory protein homolog - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C;Accession: F64075
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehn, J.C.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: F64075
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-194 <TIGR>
A;Cross-references: GB:U32736; GB:L42023; NID:g1573519; PID:g1573522; TIGR:HI0537
C;Genetics:
A;Start codon: GTG

Query Match   60.3%;   Score 38; DB 2; Length 194;
Best Local Similarity   35.7%;   Pred. No. 2.4;
Matches   5; Conservative   6; Mismatches   6; Indels   3; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           ::|| ||: :| :||
Db   64 FKLGVRLLKIFIRY 77

RESULT   5
H70474
ATP synthase F1 beta subunit - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C;Accession: H70474

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
v.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: H70474
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-478 <AQF>
A;Cross-references: GB:AE000769; NID:g2984262; PIDN:AAC07790.1; PID:g2984264; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: atpD
C;Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase
C;Keywords: P-loop
F;163-170/Region: nucleotide-binding motif A (P-loop)
F;189-358/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match   60.3%;   Score 38; DB 2; Length 478;
Best Local Similarity   50.0%;   Pred. No. 5.7;
Matches   7; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           | :|| : ||| ||
Db   369 YEVAMEVKRILQRY 382

RESULT   6
OXCKPM
acyl-CoA oxidase (EC 1.3.3.6) PXP4, peroxisomal - yeast (Candida maltosa)
C;Species: Candida maltosa
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 17-Feb-1995
C;Accession: A29441
R;Hill, D.E.; Boulay, R.; Rogers, D.
Nucleic Acids Res. 16, 365-366, 1988
A;Title: Complete nucleotide sequence of the peroxisomal acyl CoA oxidase from the al
A;Reference number: A29441; MUID:88124223
A;Accession: A29441
A;Molecule type: DNA
A;Residues: 1-709 <HI1>
A;Experimental source: ATCC 20184
C;Genetics:
A;Gene: POX4
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match   60.3%;   Score 38; DB 1; Length 709;
Best Local Similarity   50.0%;   Pred. No. 8.4;
Matches   7; Conservative   2; Mismatches   5; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           || |: | |||
Db   308 YRMLARVSTIALRY 321

RESULT   7
B69327
conserved hypothetical protein AF0618 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C;Accession: B69327
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343
A;Accession: B69327
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
A;Residues: 1-96 <KLE>
A;Cross-references: GB:AE001062; GB:AE000782; NID:g2689385; PID:g2650002; TIGR:AF0618

Query Match   57.1%; Score 36; DB 2; Length 96;
Best Local Similarity   61.5%; Pred. No. 3.1;
Matches   8; Conservative   2; Mismatches   2; Indels   3; Gaps   0;

Qy   1 YRLAIRIXRILLR 13
          ||||:|| ||:
Db   51 YRLAIKISTELLK 63

RESULT 8
OXCKAX
acyl-CoA oxidase (EC 1.3.3.6) POX4-2, peroxisomal - yeast (Candida tropicalis) (fragment
N;Alternate names: acyl-CoA oxidase II-2
C;Species: Candida tropicalis
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C;Accession: A28584
R;Small, G.M.; Lazarow, P.B.
J. Cell Biol. 105, 247-250, 1987
A;Title: Import of the carboxy-terminal portion of acyl-CoA oxidase into peroxisomes of
A;Reference number: A28584; MUID:87280361
A;Accession: A28584
A;Molecule type: mRNA
A;Residues: 1-502 <SMA>
A;Cross-references: GB:Y00623; NID:g2672; PIDN:CAA68660.1; PID:g2673
C;Genetics:
A;Gene: POX4-2
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match   57.1%; Score 36; DB 1; Length 502;
Best Local Similarity   50.0%; Pred. No. 15;
Matches   7; Conservative   2; Mismatches   5; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
          || :|: | |||
Db   101 YRMLARMSTIALRY 114

RESULT 9
OXCKX4
acyl-CoA oxidase (EC 1.3.3.6) POX4, peroxisomal - yeast (Candida tropicalis)
N;Alternate names: acyl-CoA oxidase II
C;Species: Candida tropicalis
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C;Accession: A25123
R;Okazaki, K.; Takechi, T.; Kambara, N.; Fukui, S.; Kubota, I.; Kamiryo, T.
Proc. Natl. Acad. Sci. U.S.A. 83, 1232-1236, 1986
A;Title: Two acyl-coenzyme A oxidases in peroxisomes of the yeast Candida tropicalis: pr
A;Reference number: A94084; MUID:86149279
A;Accession: A25123
A;Molecule type: DNA
A;Residues: 1-709 <OKA>
A;Cross-references: GB:M12160; NID:g170911; PIDN:AAA34362.1; PID:g170912
A;Experimental source: strain pK233, ATCC 20336
A;Comment: This enzyme, located in peroxisomes, catalyzes the oxygen-specific oxidation
; this reaction is the initial step of the peroxisomal beta-oxidation system. In C. trop
gene family.
C;Comment: The peroxisomes are eukaryote subcellular organelles that generate and degrad
ic reactions as well as the enzymes involved are usually different from the mitochondria
C;Genetics:
A;Gene: POX4
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match   57.1%; Score 36; DB 1; Length 709;
Best Local Similarity   50.0%; Pred. No. 21;

Matches   7; Conservative   2; Mismatches   5; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
          || :|: | |||
Db   308 YRMLARMSTIALRY 321

RESULT 10
OXCKX
acyl-CoA oxidase (EC 1.3.3.6) AOx, peroxisomal - yeast (Candida tropicalis)
C;Species: Candida tropicalis
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 30-Sep-1993
C;Accession: A29047
R;Murray, W.W.; Rachubinski, R.A.
Gene 51, 119-128, 1987
A;Title: The primary structure of a peroxisomal fatty acyl-CoA oxidase from the yeast
A;Reference number: A29047; MUID:87248070
A;Accession: A29047
A;Molecule type: DNA
A;Residues: 1-709 <MUR>
A;Experimental source: strain pK233, ATCC 20336
C;Genetics:
A;Gene: AOx
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match   57.1%; Score 36; DB 1; Length 709;
Best Local Similarity   50.0%; Pred. No. 21;
Matches   7; Conservative   2; Mismatches   5; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
          || :|: | |||
Db   308 YRMLARMSTIALRY 321

RESULT 11
S39842
enniatin synthetase - fungus (Fusarium scirpi)
C;Species: Fusarium scirpi
C;Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 17-Mar-1999
C;Accession: S39842; S35906; S65363
R;Haese, A.
submitted to the EMBL Data Library, November 1992
A;Reference number: S39842
A;Accession: S39842
A;Molecule type: DNA
A;Residues: 1-3131 <HAE>
A;Cross-references: EMBL:Z18755; NID:g2729; PID:g2730
R;Haese, A.; Schubert, M.; Herrmann, M.; Zocher, R.
Mol. Microbiol. 7, 905-914, 1993
A;Title: Molecular characterization of the enniatin synthetase gene encoding a multif
A;Reference number: S35906; MUID:93247491
A;Accession: S35906
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 499-1074;1572-1988;2423-2566 <HA2>
A;Cross-references: EMBL:Z18755
R;Pieper, R.; Haese, A.; Schroeder, W.; Zocher, R.
Eur. J. Biochem. 230, 119-126, 1995
A;Title: Arrangement of catalytic sites in the multifunctional enzyme enniatin synthe
A;Reference number: S65363; MUID:95324513
A;Accession: S65363
A;Molecule type: protein
A;Residues: 2029-2048;430-437;1011-1020;1021-1034;1677-1695;2294-2299 <PIE>
C;Genetics:
A;Gene: esyn1
C;Superfamily: gramicidin S synthetase I repeat homology; acetate--CoA ligase homolog
C;Keywords: multifunctional enzyme; phosphopantetheine; phosphoprotein
F;471-1082/Domain: gramicidin S synthetase I repeat homology <GRS1>
F;531-985/Domain: acetate--CoA ligase homology <ACL1>

F;1543-2574/Domain: gramicidin S synthetase I repeat homology #status atypical <GRS2>
F;1603-2044/Domain: acetate--CoA ligase homology <ACL2>
F;2507-2574/Domain: acyl carrier protein homology <ACP1>
F;2601-2667/Domain: acyl carrier protein homology <ACP2>
F;1047,2538,2632/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match   57.1%; Score 36; DB 2; Length 3131;
Best Local Similarity 53.8%; Pred. No. 87;
Matches   7; Conservative   4; Mismatches   2; Indels   0; Gaps   0;

Qy   2 RLAIRIXRILLRY 14
           ||:|| | :|::
Db 2637 KLAVRIGRRLIRH 2649

RESULT 12
OXCKP2
acyl-CoA oxidase (EC 1.3.3.6) PXP2, peroxisomal - yeast (Candida tropicalis)
C;Species: Candida tropicalis
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C;Accession: A27331
R;Okazaki, K.; Tan, H.; Fukui, S.; Kubota, I.; Kamiryo, T.
Gene 58, 37-44, 1987
A;Title: Peroxisomal acyl-coenzyme A oxidase multigene family of the yeast Candida tropi
A;Reference number: A27331; MUID:88084444
A;Accession: A27331
A;Molecule type: DNA
A;Residues: 1-724 <OKA>
A;Cross-references: GB:M18259; NID:g170909; PIDN:AAA34361.1; PID:g170910
A;Comment: This enzyme, located in peroxisomes, catalyzes the oxygen-specific oxidation
; this reaction is the initial step of the peroxisomal beta-oxidation system. In C. trop
gene family.
C;Comment: The peroxisomes are eukaryote subcellular organelles that generate and degrad
ic reactions as well as the enzymes involved are usually different from the mitochondria
C;Genetics:
A;Gene: POX2
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match   55.6%; Score 35; DB 1; Length 724;
Best Local Similarity 50.0%; Pred. No. 34;
Matches   7; Conservative   1; Mismatches   6; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           || :| |||
Db   330 YRIGARYTTIALRY 343

RESULT 13
JC4563
acyl-CoA oxidase (EC 1.3.3.6) - yeast (Candida maltosa)
C;Species: Candida maltosa
C;Date: 08-Mar-1996 #sequence_revision 19-Apr-1996 #text_change 11-Jun-1999
C;Accession: JC4563
R;Masuda, Y.; Park, S.M.; Ohta, A.; Takagi, M.
Gene 167, 157-161, 1995
A;Title: Cloning and characterization of the POX2 gene in Candida maltosa.
A;Reference number: JC4563; MUID:96144267
A;Accession: JC4563
A;Molecule type: DNA
A;Residues: 1-724 <MAS>
A;Cross-references: DDBJ:D21228; NID:g416286; PIDN:BAA04761.1; PID:d1005293; PID:g416287
A;Experimental source: P2DD
C;Comment: This protein is a member of the acyl-CoA oxidase family, and it is involved i
C;Genetics:
A;Gene: pox2
C;Superfamily: acyl-CoA oxidase
C;Keywords: oxidoreductase

Query Match   55.6%; Score 35; DB 2; Length 724;

Best Local Similarity 50.0%; Pred. No. 34;
Matches   7; Conservative   1; Mismatches   6; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           ||:|| ||| :|||
Db   330 YRIGARYTTIALRY 343

RESULT 14
F72612
hypothetical protein APE1362 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: F72612
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-157 <KAW>
A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80356.1; PID:d1044142; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1362

Query Match   54.0%; Score 34; DB 2; Length 157;
Best Local Similarity 58.3%; Pred. No. 12;
Matches   7; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   3 LAIRIXRILLRY 14
           || :|| ||| :|
Db   35 LAAKIARILTKY 46

RESULT 15
F72721
probable citrate lyase beta chain APE0311 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: F72721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KAW>
A;Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79266.1; PID:d1043052; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0311

Query Match   54.0%; Score 34; DB 2; Length 276;
Best Local Similarity 50.0%; Pred. No. 21;
Matches   7; Conservative   2; Mismatches   5; Indels   0; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
           |||:| ||| :|
Db   259 YRLALNLLRRASRY 272

Search completed: February 7, 2000, 11:54:32
Job time: 2342 sec

```
                                                                                      P20530 vaccinia v1
                                                                                      P28544 chlamydia t
                                                                                      O59300 pyrococcus
                                                                                      P22654 bovine coro
                                                                                      P10525 bovine coro
                                                                                      P26626 turkey ente
                                                                                      P52133 escherichia
                                                                                      Q09465 caenorhabdi
                                                                                      P77365 escherichia
                                                                                      P72980 synechocyst
                                                                                      P52024 pseudomonas
```

```
35  30  47.6  72  1  YVAU_VACCC
36  30  47.6 133  1  RS8_CHLTR
37  30  47.6 142  1  RL13_PYRHO
38  30  47.6 207  1  YIOR_CVBF
39  30  47.6 207  1  YIOR_CVBM
40  30  47.6 207  1  YIOR_CVTKE
41  30  47.6 233  1  YFJR_ECOLI
42  30  47.6 274  1  YQ59_CAEEL
43  30  47.6 285  1  YAFY_ECOLI
44  30  47.6 289  1  LIP1_SYNY3
45  30  47.6 326  1  HOLB_PSEAE
```

ALIGNMENTS

```
RESULT  1
UREF_ACTPL
ID   UREF_ACTPL       STANDARD;      PRT;     227 AA.
AC   O54423;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   UREASE ACCESSORY PROTEIN UREF.
GN   UREF.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CM5;
RA   BOSSE J.T., MACINNES J.I.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
-!- SIMILARITY: BELONGS TO THE UREF FAMILY.
-------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-------------------------------------------------------------------------
DR   EMBL; U89957; AAC00063.1; -.
KW   Nickel.
SQ   SEQUENCE   227 AA;   25397 MW;  310CB946 CRC32;
```

```
Query Match             66.7%;  Score 42;  DB 1;  Length 227;
Best Local Similarity   42.9%;  Pred. No. 0.16;
Matches    6;  Conservative   5;  Mismatches  3;  Indels   0;  Gaps   0;

Qy        1 YRLAIRIXRILLRY 14
            |:|.:|.:|.:|:||
Db       97 YKLGVRLLKIFIRY 110
```

```
RESULT  2
ATPB_MYCLE
ID   ATPB_MYCLE       STANDARD;      PRT;     485 AA.
AC   P45823;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN   ATPD.
OS   Mycobacterium leprae.
OC   Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMITH D.R., ROBISON K.;
```

```
                   GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        February  8, 2000, 00:59:58 ; Search time 63.71 seconds
                                             (without alignments)
                                             6.563 Million cell updates/sec

Title:          US-08-653-294-20
Perfect score:  63
Sequence:       1 YRLAIRIXRILLRY 14

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters:       82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :      SwissProt_38:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.

                          SUMMARIES
```

```
                   %
Result       Query
No.   Score  Match  Length DB  ID          Description
----  -----  -----  ------ --  ----------  ------------------
   1     42   66.7    227  1   UREF_ACTPL  O54423 actinobacil
   2     39   61.9    485  1   ATPB_MYCLE  P45823 mycobacteri
   3     39   61.9    486  1   ATPB_MYCTU  Q10593 mycobacteri
   4     38   60.3    194  1   UREF_HAEIN  P44395 haemophilus
   5     38   60.3    471  1   ATPB_HERAU  P42466 herpetosiph
   6     38   60.3    478  1   ATPB_AQUAE  O67828 aquifex aeo
   7     38   60.3    478  1   ATPB_AQUPY  O50292 aquifex pyr
   8     38   60.3    708  1   CAO4_CANMA  P05335 candida mal
   9     36   57.1    502  1   CAO3_CANTR  P11355 candida tro
  10     36   57.1    708  1   CAO2_CANTR  P06598 candida tro
  11     35   55.6    723  1   CAO4_CANTR  P11356 candida tro
  12     35   55.6    724  1   CAO2_CANMA  Q00468 candida mal
  13     34   54.0    323  1   ACCA_AQUAE  O67260 aquifex aeo
  14     34   54.0    446  1   NY5R_CANFA  O62729 canis famil
  15     34   54.0    446  1   NY5R_PIG    O97969 sus scrofa
  16     34   54.0    455  1   NY5R_HUMAN  Q15761 homo sapien
  17     34   54.0    456  1   NY5R_RAT    Q63634 rattus norv
  18     34   54.0    466  1   NY5R_MOUSE  O70342 mus musculu
  19     34   54.0    661  1   CAO1_CANTR  P08790 candida tro
  20     34   54.0   2009  1   SEC7_YEAST  P11075 saccharomyc
  21     33   52.4    260  1   AHYR_AERHY  Q44059 aeromonas h
  22     33   52.4    470  1   ATPB_LACCA  Q03234 lactobacill
  23     33   52.4    477  1   ATPB_STRLI  P50004 streptomyce
  24     33   52.4   2376  1   YIM9_YEAST  P40468 saccharomyc
  25   32.5   51.6   2569  1   LMA3_MOUSE  Q61789 mus musculu
  26     32   50.8    205  1   FGFH_NPVOP  O10284 orgyia pseu
  27     32   50.8    468  1   ATPB_ENTHR  P43451 enterococcu
  28     31   49.2    112  1   RL31_YEAST  P04649 saccharomyc
  29     31   49.2    195  1   PACA_CLAMA  P48144 clarias mac
  30     31   49.2    343  1   YSC2_THEFL  P25125 thermus aqu
  31     31   49.2    626  1   YT17_MYCTU  Q10966 mycobacteri
  32     31   49.2   1603  1   VIT5_CAEEL  P06125 caenorhabdi
  33     31   49.2   3433  1   UTRO_HUMAN  P46939 homo sapien
  34     30   47.6     57  1   YCU5_CAEEL  Q22702 caenorhabdi
```

```
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U15186; AAA63108.1; -.
DR   HSSP; P07677; 1SKY.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR   PFAM; PF00006; ATP-synt_ab; 1.
DR   PFAM; PF00306; ATP-synt_ab_C; 1.
KW   Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW   Hydrogen ion transport.
FT   NP_BIND      170    177      ATP (POTENTIAL).
SQ   SEQUENCE    485 AA;  53034 MW;  07216783 CRC32;

Query Match              61.9%;  Score 39;  DB 1;  Length 485;
Best Local Similarity    57.1%;  Pred. No. 1.5;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 YRLAIRIXRILLRY 14
             |||:| .||||:||
Db       376 YRVAQEVIRILQRY 389

RESULT 3
ATPB_MYCTU
ID   ATPB_MYCTU     STANDARD;      PRT;   486 AA.
AC   Q10593;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN   ATPD OR RV1310 OR MTCY373.30.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.


CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; Z73419; CAA97743.1; -.
DR   HSSP; P07677; 1SKY.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR   PFAM; PF00006; ATP-synt_ab; 1.
DR   PFAM; PF00306; ATP-synt_ab_C; 1.
KW   Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW   Hydrogen ion transport.
FT   NP_BIND      171    178      ATP (POTENTIAL).
SQ   SEQUENCE    486 AA;  53094 MW;  A8001B2F CRC32;

Query Match              61.9%;  Score 39;  DB 1;  Length 486;
Best Local Similarity    57.1%;  Pred. No. 1.5;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 YRLAIRIXRILLRY 14
             |||:| .||||:||
Db       377 YRVAQEVIRILQRY 390

RESULT 4
UREF_HAEIN
ID   UREF_HAEIN     STANDARD;      PRT;   194 AA.
AC   P44395;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   UREASE ACCESSORY PROTEIN UREF.
GN   UREF OR HI0537.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
CC   -!- SIMILARITY: BELONGS TO THE UREF FAMILY.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U32736; AAC22195.1; -.
DR   TIGR; HI0537; -.
KW   Nickel.
SQ   SEQUENCE    194 AA;  21960 MW;  44EB0D8E CRC32;
```

```
Query Match          60.3%;          Score 38;   DB 1;   Length 194;
Best Local Similarity   35.7%;        Pred. No. 0.9;
Matches   5;   Conservative   3;   Mismatches   6;   Indels   0;   Gaps   0;

Qy     1 YRLAIRIXRILLRY 14
           ::|::|: :|:|||
Db    64 FKLGVRLLKIFIRY 77
```

RESULT 5
ATPB_HERAU
```
ID  ATPB_HERAU       STANDARD;       PRT;    471 AA.
AC  P42466;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  01-NOV-1995 (Rel. 32, Last annotation update)
DE  ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN  ATPD.
OS  Herpetosiphon aurantiacus (Herpetosiphon giganteus).
OC  Bacteria; Green non-sulfur bacteria; Chloroflexaceae group;
OC  Herpetosiphon.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=HPGA1;
RX  MEDLINE; 94368062.
RA  LUDWIG W., NEUMAIER J., KLUGBAUER N., BROCKMANN E., ROLLER C.,
RA  KLUGBAUER S., REETZ K., SCHACHTNER I., LUDVIGSEN A.,
RA  BACHLEITNER M., FISCHER U., SCHLEIFER K.H.;
RT  "Phylogenetic relationships of Bacteria based on comparative sequence
RT  analysis of elongation factor Tu and ATP-synthase beta-subunit
RT  genes.";
RL  Antonie Van Leeuwenhoek 64:285-305(1993).
CC  -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC      GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC      SUBUNIT.
CC  -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC      CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC      SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC  -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC  --------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------
DR  EMBL; X76876; CAA54203.1; -.
DR  HSSP; P07677; 1SKY.
DR  PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR  PFAM; PF00006; ATP-synt_ab; 1.
DR  PFAM; PF00306; ATP-synt_ab_C; 1.
KW  Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW  Hydrogen ion transport.
FT  NP_BIND   152    159      ATP (BY SIMILARITY).
SQ  SEQUENCE   471 AA;  51057 MW;  AAE39561 CRC32;
```

```
Query Match          60.3%;          Score 38;   DB 1;   Length 471;
Best Local Similarity   50.0%;        Pred. No. 2.3;
Matches   7;   Conservative   3;   Mismatches   3;   Indels   4;   Gaps   0;

Qy     1 YRLAIRIXRILLRY 14
           ||:| : |:| ||
Db   364 YRVATEVQRMLQRY 377
```

RESULT 6
ATPB_AQUAE
```
ID  ATPB_AQUAE       STANDARD;       PRT;    478 AA.
AC  O67828;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1999 (Rel. 39, Last annotation update)
DE  ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN  ATPD OR AQ_2038.
OS  Aquifex aeolicus.
OC  Bacteria; Aquificales; Aquificaceae; Aquifex.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=VF5;
RX  MEDLINE; 98196666.
RA  DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA  GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA  FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus.";
RL  Nature 392:353-358(1998).
CC  -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC      GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC      SUBUNIT.
CC  -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC      CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC      SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC  -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC  --------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------
DR  EMBL; AE000769; AAC07790.1; -.
DR  PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR  PFAM; PF00006; ATP-synt_ab; 1.
DR  PFAM; PF00306; ATP-synt_ab_C; 1.
KW  Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW  Hydrogen ion transport.
FT  NP_BIND   163    170      ATP (POTENTIAL).
SQ  SEQUENCE   478 AA;  53321 MW;  0E23DB68 CRC32;
```

```
Query Match          60.3%;          Score 38;   DB 1;   Length 478;
Best Local Similarity   50.0%;        Pred. No. 2.4;
Matches   7;   Conservative   3;   Mismatches   4;   Indels   0;   Gaps   0;

Qy     1 YRLAIRIXRILLRY 14
           |:|:: |||:||
Db   369 YEVAMEVKRILQRY 382
```

RESULT 7
ATPB_AQUPY
```
ID  ATPB_AQUPY       STANDARD;       PRT;    478 AA.
AC  O50292;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN  ATPD.
OS  Aquifex pyrophilus.
OC  Bacteria; Aquificales; Aquificaceae; Aquifex.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=DSM 6858;
RX  MEDLINE; 98248216.
RA  LUDWIG W., STRUNK O., KLUGBAUER S., KLUGBAUER N., WEIZENEGGER M.,
RA  NEUMAIER J., BACHLEITNER M., SCHLEIFER K.H.;
RT  "Bacterial phylogeny based on comparative sequence analysis.";
RL  Electrophoresis 19:554-568(1998).
```

```
CC  -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC      GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC      SUBUNIT.
CC  -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC      CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC      SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC  -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC  ------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ------------------------------------------------------------------------
DR  EMBL; Y15786; CAA75780.1; -.
DR  PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR  PFAM; PF00006; ATP-synt_ab; 1.
DR  PFAM; PF00306; ATP-synt_ab_C; 1.
KW  Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW  Hydrogen ion transport.
FT  NP_BIND     163    170    ATP (POTENTIAL).
SQ  SEQUENCE    478 AA;    53393 MW;    8F8ADBE9 CRC32;

Query Match             60.3%;  Score 38;  DB 1;  Length 478;
Best Local Similarity   50.0%;  Pred. No. 2.4;
Matches    7;  Conservative   3;  Mismatches   4;  Indels   0;  Gaps   0;

Qy      1  YRLAIRIXRILLRY 14
           | :|: :|||||||
Db    369  YEVAMEVKRILQRY 382

RESULT 8
CAO4_CANMA
ID  CAO4_CANMA    STANDARD;    PRT;    708 AA.
AC  P05335;
DT  01-NOV-1988 (Rel. 09, Created)
DT  01-MAY-1992 (Rel. 22, Last sequence update)
DT  01-MAY-1992 (Rel. 22, Last annotation update)
DE  ACYL-COENZYME A OXIDASE POX4 (EC 1.3.6) (ACYL-COA OXIDASE) (AOX).
GN  POX4.
OS  Candida maltosa (Yeast).
OC  Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC  Candidaceae; Candida.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 20184;
RX  MEDLINE; 88124223.
RA  HILL D.E., BOULAY R., ROGERS D.;
RT  "Complete nucleotide sequence of the peroxisomal acyl CoA oxidase
RT  from the alkane-utilizing yeast Candida maltosa.";
RL  Nucleic Acids Res. 16:365-366(1988).
CC  -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC      H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC      FROM 8 TO 18).
CC  -!- COFACTOR: FAD.
CC  -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC      SYSTEM.
CC  -!- SUBUNIT: HOMOOCTAMER.
CC  -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC  ------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ------------------------------------------------------------------------


DR  EMBL; X06721; CAA29901.1; -.
DR  PIR; A29441; OXCKPM.
KW  Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW  Multigene family.
FT  INIT_MET    0    0    BY SIMILARITY.
SQ  SEQUENCE    708 AA;    78242 MW;    D5E344D2 CRC32;

Query Match             60.3%;  Score 38;  DB 1;  Length 708;
Best Local Similarity   50.0%;  Pred. No. 3.6;
Matches    7;  Conservative   2;  Mismatches   0;  Indels   5;  Gaps   0;

Qy      1  YRLAIRIXRILLRY 14
           ||: |: | |||
Db    307  YRMLARVSTIALRY 320

RESULT 9
CAO3_CANTR
ID  CAO3_CANTR    STANDARD;    PRT;    502 AA.
AC  P11355;
DT  01-JUL-1989 (Rel. 11, Created)
DT  01-JUL-1989 (Rel. 11, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  ACYL-COENZYME A OXIDASE POX4-2 (EC 1.3.6) (ACYL-COA OXIDASE)
DE  (FRAGMENT).
GN  POX4-2.
OS  Candida tropicalis (Yeast).
OC  Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC  Candidaceae; Candida.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 87280361.
RA  SMALL G.M., LAZAROW P.B.;
RT  "Import of the carboxy-terminal portion of acyl-CoA oxidase into
RT  peroxisomes of Candida tropicalis.";
RL  J. Cell Biol. 105:247-250(1987).
CC  -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC      H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC      FROM 8 TO 18).
CC  -!- COFACTOR: FAD.
CC  -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC      SYSTEM.
CC  -!- SUBUNIT: HOMOOCTAMER.
CC  -!- SUBCELLULAR LOCATION: PEROXISOMAL.
DR  PIR; A28584; OXCKAX.
KW  Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW  Multigene family.
FT  NON_TER     1    1
SQ  SEQUENCE    502 AA;    55528 MW;    EACE80C4 CRC32;

Query Match             57.1%;  Score 36;  DB 1;  Length 502;
Best Local Similarity   50.0%;  Pred. No. 6.4;
Matches    7;  Conservative   2;  Mismatches   2;  Indels   5;  Gaps   0;

Qy      1  YRLAIRIXRILLRY 14
           ||: |: :| |||
Db    101  YRMLARMSTIALRY 114

RESULT 10
CAO2_CANTR
ID  CAO2_CANTR    STANDARD;    PRT;    708 AA.
AC  P06598;
DT  01-JAN-1988 (Rel. 06, Created)
DT  01-NOV-1988 (Rel. 09, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  ACYL-COENZYME A OXIDASE II (EC 1.3.6) (ACYL-COA OXIDASE) (PXP-4).
GN  AOX OR POX-4.
OS  Candida tropicalis (Yeast).
OC  Eukaryota: Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC  Candidaceae; Candida.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20336 / PK233;
RX   MEDLINE; 87248070.
RA   MURRAY W.W., RACHUBINSKI R.A.;
RT   "The primary structure of a peroxisomal fatty acyl-CoA oxidase from
RT   the yeast Candida tropicalis pk233.";
RL   Gene 51:119-128(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20336 / PK233;
RX   MEDLINE; 86149279.
RA   OKAZAKI K., TAKECHI T., KAMBARA N., FUKUI S., KUBOTA I., KAMIRYO T.;
RT   "Two acyl-coenzyme A oxidases in peroxisomes of the yeast Candida
RT   tropicalis: primary structures deduced from genomic DNA sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1232-1236(1986).
RN   [3]
RP   SEQUENCE OF 208-709 FROM N.A.
RC   STRAIN=RR1;
RX   MEDLINE; 87280361.
RA   SMALL G.M., LAZAROW P.B.;
RT   "Import of the carboxy-terminal portion of acyl-CoA oxidase into
RT   peroxisomes of Candida tropicalis.";
RL   J. Cell Biol. 105:247-250(1987).
CC   -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC       H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC       FROM 8 TO 18).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC       SYSTEM.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M16193; AAA34322.1; -.
DR   EMBL; M12160; AAA34362.1; -.
DR   EMBL; Y00623; CAA68660.1; -.
DR   EMBL; Y00623; CAA68661.1; ALT_INIT.
DR   EMBL; Y00623; CAA68662.1; ALT_INIT.
DR   PIR; A25123; OXCKX4.
DR   PIR; A29047; OXCKX.
KW   Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW   Multigene family.
FT   INIT_MET       0      0
FT   CONFLICT     216    216      Q -> E (IN REF. 3).
FT   CONFLICT     245    245      P -> A (IN REF. 2 AND 3).
FT   CONFLICT     335    335      N -> K (IN REF. 2).
FT   CONFLICT     393    393      FRAMESHIFT ERROR (IN REF. 2 AND 3).
FT   CONFLICT     435    435      H -> Y (IN REF. 2).
FT   CONFLICT     462    462      G -> A (IN REF. 2).
FT   CONFLICT     495    495      E -> S (IN REF. 2).
FT   CONFLICT     576    578      ELA -> DLV (IN REF. 2);
FT   CONFLICT     697    697      Q -> E (IN REF. 2 AND 3).
SQ   SEQUENCE   708 AA;  79041 MW;  D97A4EC8 CRC32;
```

```
Query Match             57.1%;
Best Local Similarity   50.0%;    Score 36;  DB 1;  Length 708;
Matches    7;  Conservative  2;  Mismatches  5;  Indels  0;  Gaps  0;

Qy      1 YRLAIRIXRILLRY 14
          |: |.; | |||
Db    307 YRMLARMSTIALRY 320

RESULT 11
```

```
CAO4_CANTR
ID   CAO4_CANTR     STANDARD;      PRT;    723 AA.
AC   P11356;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ACYL-COENZYME A OXIDASE PXP-2 (EC 1.3.6) (ACYL-COA OXIDASE).
GN   POX2.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Candidaceae; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88084444.
RA   OKAZAKI K., TAN H., FUKUI S., KUBOTA I., KAMIRYO T.;
RT   "Peroxisomal acyl-coenzyme A oxidase multigene family of the yeast
RT   Candida tropicalis: nucleotide sequence of a third gene and its
RT   protein product.";
RL   Gene 58:37-44(1987).
CC   -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC       H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC       FROM 8 TO 18).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC       SYSTEM.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M18259; AAA34361.1; -.
DR   PIR; A27331; OXCKP2.
KW   Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW   Multigene family.
FT   INIT_MET       0      0
SQ   SEQUENCE   723 AA;  81804 MW;  6DC2D2B7 CRC32;
```

```
Query Match             55.6%;    Score 35;  DB 1;  Length 723;
Best Local Similarity   50.0%;    Pred. No. 15;
Matches    7;  Conservative  1;  Mismatches  1;  Indels  6;  Gaps  0;

Qy      1 YRLAIRIXRILLRY 14
          |: |.;  | |||
Db    329 YRIGARYTTIALRY 342

RESULT 12
CAO2_CANMA
ID   CAO2_CANMA     STANDARD;      PRT;    724 AA.
AC   Q00468;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ACYL-COENZYME A OXIDASE POX2 (EC 1.3.6) (ACYL-COA OXIDASE) (AOX).
GN   POX2.
OS   Candida maltosa (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Candidaceae; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IAM 12247;
RX   MEDLINE; 96144267.
RA   MASUDA Y., PARK S.M., OHTA A., TAKAGI M.;
RT   "Cloning and characterization of the POX2 gene in Candida maltosa.";
RL   Gene 167:157-161(1995).
CC   -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
```

```
CC       H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC       FROM 8 TO 18).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC       SYSTEM.
CC   -!- SUBUNIT: HOMOOCTAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D21228; BAA04761.1; -.
KW   Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW   Multigene family.
SQ   SEQUENCE   724 AA;   82273 MW;   1AE92F21 CRC32;

Query Match             55.6%;  Score 35;  DB 1;  Length 724;
Best Local Similarity   50.0%;  Pred. No. 15;
Matches    7;  Conservative   1;  Mismatches   6;  Indels   0;  Gaps   0;

Qy     1  YRLAIRIXRILLRY 14
          ||:: | | |||
Db   330  YRIGARYTTIALRY 343


RESULT   13
ACCA_AQUAE
ID   ACCA_AQUAE     STANDARD;     PRT;     323 AA.
AC   O67260;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT ALPHA
DE   (EC 6.4.1.2).
GN   ACCA OR AQ_1206.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
CC       CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE
CC       CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
CC       TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CARBOXYBIOTIN CARBOXYL CARRIER PROTEIN +
CC       ACETYL-COA = BIOTIN CARBOXYL CARRIER PROTEIN + MALONYL-COA.
CC   -!- PATHWAY: FIRST STEP IN LONG-CHAIN FATTY ACID SYNTHESIS.
CC   -!- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN
CC       CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS
CC       OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
CC   -!- SIMILARITY: TO THE C-TERMINUS OF MAMMALIAN PROPIONYL-COA
CC       CARBOXYLASE BETA CHAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE000728; AAC07216.1; -.
KW   Fatty acid biosynthesis; Ligase.
SQ   SEQUENCE   323 AA;   36198 MW;   4A96C81E CRC32;

Query Match             54.0%;  Score 34;  DB 1;  Length 323;
Best Local Similarity   42.9%;  Pred. No. 10;
Matches    6;  Conservative   4;  Mismatches   4;  Indels   0;  Gaps   0;

Qy     1  YRLAIRIXRILLRY 14
          || |:: ||
Db   140  YRKAIKVFKLAERY 153


RESULT   14
NY5R_CANFA
ID   NY5R_CANFA     STANDARD;     PRT;     446 AA.
AC   O62729;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   NEUROPEPTIDE Y RECEPTOR TYPE 5 (NPY5-R) (NPY-Y5 RECEPTOR) (Y5
DE   RECEPTOR) (NPY5).
GN   NPY5R OR NPY5.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99017379.
RA   BOROWSKY B., WALKER M.W., BARD J., WEINSHANK R.L., LAZ T.M.,
RA   VAYSSE P., BRANCHEK T.A., GERALD C.;
RT   "Molecular biology and pharmacology of multiple NPY Y5 receptor
RT   species homologs.";
RL   Regul. Pept. 75:45-53(1998).
CC   -!- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
CC       CYCLASE ACTIVITY. SEEMS TO BE ASSOCIATED WITH FOOD INTAKE. COULD
CC       BE INVOLVED IN FEEDING DISORDERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- HIGHEST TO TACHYKININS RECEPTORS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF049328; AAC17838.1; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR   PFAM; PF00001; 7tm_1; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Lipoprotein; Palmitate.
FT   DOMAIN          1     39       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM       41     62       1 (POTENTIAL).
FT   DOMAIN         63     74       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM       75     95       2 (POTENTIAL).
FT   DOMAIN         96    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      116    137       3 (POTENTIAL).
FT   DOMAIN        138    157       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      158    178       4 (POTENTIAL).
FT   DOMAIN        179    210       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      211    232       5 (POTENTIAL).
FT   DOMAIN        233    311       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      369    391       6 (POTENTIAL).
FT   DOMAIN        392    404       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      405    428       7 (POTENTIAL).
FT   DOMAIN        429    446       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD       10     10       POTENTIAL.
```

```
FT   CARBOHYD      17     17       POTENTIAL.
FT   DISULFID     114    198       BY SIMILARITY.
FT   LIPID        442    442       PALMITATE (POTENTIAL).
SQ   SEQUENCE     446 AA;  51012 MW;  7D2CD74A CRC32;

Query Match              54.0%;  Score 34;  DB 1;  Length 446;
Best Local Similarity    42.9%;  Pred. No. 14;
Matches   6;  Conservative   4;  Mismatches   4;  Indels   0;  Gaps   0;

Qy      1 YRLAIRIXRILLRY 14
          ||:|  |  :||:::|
Db    207 YRIAFTISLLLVQY 220


RESULT 15
NY5R_PIG
ID   NY5R_PIG       STANDARD;       PRT;    446 AA.
AC   O97969;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   NEUROPEPTIDE Y RECEPTOR TYPE 5 (NPY5-R) (NPY-Y5 RECEPTOR) (Y5
DE   RECEPTOR).
GN   NPY5R OR NPYR5.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WRAITH A., TORNSTEN A., CHARDON P., HARBITZ I., CHOWDHARY B.P.,
RA   ANDERSSON L., LARHAMMAR D.;
RT   "Porcine NPY receptors NPY1R, NPY2R and NPY5R: cloning, mapping and
RT   comparative analysis."; Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LWD; TISSUE=KIDNEY;
RA   ITO Y., MINEZAWA M.;
RT   "Sus scrofa gene for neuropeptide Y receptor type 5, complete cds.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
CC       CYCLASE ACTIVITY. SEEMS TO BE ASSOCIATED WITH FOOD INTAKE. COULD
CC       BE INVOLVED IN FEEDING DISORDERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       HIGHEST TO TACHYKININS RECEPTORS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF106083; AAD13778.1; -.
DR   EMBL; AB019185; BAA34055.1; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Lipoprotein; Palmitate.
FT   DOMAIN         1     39       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      41     62       1 (POTENTIAL).
FT   DOMAIN        63     74       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      75     95       2 (POTENTIAL).
FT   DOMAIN        96    115       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     116    137       3 (POTENTIAL).
FT   DOMAIN       138    157       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     158    178       4 (POTENTIAL).
FT   DOMAIN       179    210       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     211    232       5 (POTENTIAL).
FT   DOMAIN       233    311       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     369    391       6 (POTENTIAL).
FT   DOMAIN       392    404       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     405    428       7 (POTENTIAL).
FT   DOMAIN       429    446       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      10     10       POTENTIAL.
FT   CARBOHYD      17     17       POTENTIAL.
FT   DISULFID     114    198       BY SIMILARITY.
FT   LIPID        442    442       PALMITATE (POTENTIAL).
SQ   SEQUENCE     446 AA;  50474 MW;  79A4E2F3 CRC32;

Query Match              54.0%;  Score 34;  DB 1;  Length 446;
Best Local Similarity    42.9%;  Pred. No. 14;
Matches   6;  Conservative   4;  Mismatches   4;  Indels   0;  Gaps   0;

Qy      1 YRLAIRIXRILLRY 14
          ||:|  |  :||:::|
Db    207 YRIAFTISLLLVQY 220


Search completed: February  8, 2000, 00:59:58
Job time: 3787 sec
```

THIS PAGE BLANK (USPTO)

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        February  8, 2000, 13:17:48 ; Search time 209.03 Seconds
                                              (without alignments)
                                              4.644 Million cell updates/sec

Title:          US-08-653-294-20
Perfect score:  63
Sequence:       1 YRLAIRIXRILLRY 14

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters:   225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :      SPTREMBL_12:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	60.3	133	2	Q92903	Q92903 chlamydia p
2	38	60.3	2810	5	Q20456	Q20456 caenorhabdi
3	37	58.7	1592	5	O45019	O45019 caenorhabdi
4	36	57.1	96	1	O29637	O29637 archaeoglob
5	36	57.1	248	3	O74720	O74720 debaryomyce
6	36	57.1	353	2	O70073	O70073 agrobacteri
7	36	57.1	3131	3	O00869	O00869 fusarium sc
8	35	55.6	265	2	Q44541	Q44541 azotobacter
9	35	55.6	402	2	O57090	O57090 corynebacte
10	35	55.6	404	10	Q9XIL2	Q9XIL2 arabidopsis
11	35	55.6	469	2	O50159	O50159 streptococc
12	34	54.0	44	10	O49025	O49025 gracillaria
13	34	54.0	157	1	Q9YC92	Q9YC92 aeropyrum p
14	34	54.0	170	2	P94457	P94457 bacillus st
15	34	54.0	276	1	Q9YFD0	Q9YFD0 aeropyrum p
16	34	54.0	372	13	O93237	O93237 cyprinus ca
17	34	54.0	468	2	Q9ZJ01	Q9ZJ01 streptococc
18	34	54.0	1021	4	O15451	O15451 homo sapien
19	34	54.0	1251	4	O15450	O15450 homo sapien
20	33	52.4	149	2	Q9XIG5	Q9XIG5 thermotoga

21	33	52.4	485	2	P71815	P71815 mycobacteri
22	33	52.4	499	10	O04376	O04376 arabidopsis
23	33	52.4	532	2	P72007	P72007 mycobacteri
24	33	52.4	719	3	Q9I7B1	Q9I7B1 pichia past
25	33	52.4	2630	12	O55319	O55319 acyrthosiph
26	33	52.4	3074	12	Q08534	Q08534 sugar beet
27	32.5	51.6	42	11	O54741	O54741 mus musculu
28	32	50.8	339	12	Q9YA89	Q9ya89 aeropyrum p
29	32	50.8	422	12	Q88530	Q88530 turkey herp
30	32	50.8	591	2	Q928G2	Q928g2 chlamydia p
31	32	50.8	785	10	O23072	O23072 arabidopsis
32	32	50.8	2962	5	Q93326	Q93326 caenorhabdi
33	32	50.8	3070	12	O89906	O89906 beet yellow
34	31	49.2	76	3	O06739	O06739 saccharomyc
35	31	49.2	169	11	O70180	O70180 rattus norv
36	31	49.2	185	2	O55192	O55192 synechocyst
37	31	49.2	192	3	O13610	O13610 schizosacch
38	31	49.2	211	5	Q9XU58	Q9XU58 caenorhabdi
39	31	49.2	220	1	Q9Y961	Q9Y961 aeropyrum p
40	31	49.2	223	2	Q9ZIN6	Q9ZIN6 staphylococ
41	31	49.2	224	2	O83422	O83422 treponema p
42	31	49.2	275	5	O93780	O93780 caenorhabdi
43	31	49.2	277	2	Q52027	Q52027 pseudomonas
44	31	49.2	294	8	O02676	O02676 podospora a
45	31	49.2	331	1	O58520	O58520 Pyrococcus

ALIGNMENTS

```
RESULT   1
Q92903
ID   Q92903                PRELIMINARY;       PRT;    133 AA.
AC   Q92903;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 16.1 KD PROTEIN.
GN   CPN0181.
OS   Chlamydia pneumoniae.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RA   KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA   GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT   "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001604; AAD18334.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE  133 AA;  16132 MW;  8DA54C6A CRC32;

Query Match              60.3%;   Score 38;  DB 2;   Length 133;
Best Local Similarity    61.5%;   Pred. No. 4.3;
Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 RLAIRIXRILLRY 14
           |: |: |||||||
Db      88 RIPWRLKRILLRY 100

RESULT   2
Q20456
ID   Q20456                PRELIMINARY;       PRT;    2810 AA.
AC   Q20456;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HUM-4 PROTEIN.
GN   HUM-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
```

```
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   COTTAGE A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94157518.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z66563; CAA91469.1; -.
DR   PFAM; PF00612; IQ; 2.
DR   PFAM; PF00063; myosin_head; 4.
DR   PFAM; PF00784; MyTH4; 2.
SQ   SEQUENCE 2810 AA; 323526 MW; 6274286C CRC32;

Query Match       60.3%;  Score 38;  DB 5;  Length 2810;
Best Local Similarity  35.7%;  Pred. No. 86;
Matches    5;  Conservative 7;  Mismatches 2;  Indels 2;  Gaps 0;

Qy      1 YRLAIRIXRILLRY 14
          :|| :::|:|:|
Db   1102 FRLSVEIFKLILKY 1115

RESULT 3
ID   O45019      PRELIMINARY;      PRT;   1592 AA.
AC   O45019;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ZC123.3 PROTEIN.
GN   ZC123.3
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   BRADSHAW H., GRAVES T., BIEWALD T.;
RT   "The sequence of C. elegans cosmid ZC123.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RA   NONE.
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AF043706; AAB97603.2; -.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 6.
DR   PFAM; PF00096; zf-C2H2; 2.

KW   Homeobox; DNA-binding; Nuclear protein; zinc-finger; Metal-binding.
SQ   SEQUENCE 1592 AA;  178053 MW;  BAFDE8CE CRC32;

Query Match       58.7%;  Score 37;  DB 5;  Length 1592;
Best Local Similarity  72.7%;  Pred. No. 77;
Matches    8;  Conservative 2;  Mismatches 1;  Indels 0;  Gaps 0;

Qy      2 RLAIRIXRILL 12
          |:||:| ||||
Db    477 RVAIRLMRILL 487

RESULT 4
ID   O29637      PRELIMINARY;      PRT;   96 AA.
AC   O29637;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   AF0618.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001062; AAB90622.1; -.
DR   TIGR; AF0618; -.
KW   Hypothetical protein.
SQ   SEQUENCE 96 AA;  10892 MW;  76C3565A CRC32;

Query Match       57.1%;  Score 36;  DB 1;  Length 96;
Best Local Similarity  61.5%;  Pred. No. 7.6;
Matches    8;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0;

Qy      1 YRLAIRIXRILLR 13
          |||||:| |:
Db     51 YRLAIKISTELLK 63

RESULT 5
ID   O74720      PRELIMINARY;      PRT;   248 AA.
AC   O74720;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   LINEAR PLASMID PDHL1, ORF1 AND ORF2, PARTIAL (FRAGMENT).
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OG   Plasmid pDHL1.
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Debaryomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC90624, CBS7848;
```

RA GUNGE N.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC90624, CBS7848;
RX MEDLINE; 97344364.
RA FUKUDA K., MAEBUCHI M., TAKATA H., GUNGE N.;
RT "The linear plasmid pDHL1 from Debaryomyces hansenii encodes a protein
RT highly homologous to the pGKl1-plasmid DNA polymerase.";
RL Yeast 13:613-620(1997).
DR EMBL; AJ011124; CAA09498.1; -.
KW Plasmid.
FT NON_TER 248 248
SQ SEQUENCE 248 AA; 27239 MW; 54CD45A4 CRC32;

Query Match
Best Local Similarity 57.1%; Score 36; DB 3; Length 248;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YRLAIRIXRILL 12
         ||:|| |||
Db 4 YRLAIFIASILL 15

RESULT 6
ID O70073
AC O70073; PRELIMINARY; PRT; 353 AA.
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE PUTATIVE PERIPLASMIC PROTEIN CHTH.
GN CHTH.
OS Agrobacterium tumefaciens, and Agrobacterium sp.
OG Plasmid pTiChry5.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHRY5;
RA OGER P.M., VAUDEQUIN V., DESSAUX Y.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ANT4;
RA OGER P.M., VAUDEQUIN-DRANSART V., DESSAUX Y.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF065242; AAC17432.1; -.
DR EMBL; U67851; AAC12800.1; -.
KW Plasmid.
SQ SEQUENCE 353 AA; 38687 MW; 4B019D49 CRC32;

Query Match
Best Local Similarity 57.1%; Score 36; DB 2; Length 353;
Matches 6; Conservative 6; Mismatches 4; Indels 3; Gaps 0;

Qy 2 RLAIRIXRILLRY 14
        :|:|: |||
Db 247 KLGLRLGRVRLRY 259

RESULT 7
ID Q00869
AC Q00869; PRELIMINARY; PRT; 3131 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ENNIATIN SYTHETASE.
GN ESYN1.
OS Fusarium scirpi.

OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC Hypocreales; Hypocreaceae; anamorphic Hypocreaceae; Fusarium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LAMBOTTE ET FAUTREY;
RA HAESE A., SCHUBERT M., HERRMANN M., ZOCHER R.;
RL Mol. Microbiol. 0:0-0(1992).
DR EMBL; Z18755; CAA79245.1; -.
DR MENDEL; 20784; Fussr;3105;20784.
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PFAM; PF00501; AMP-binding; 2.
DR PFAM; PF00668; DUF4; 2.
DR PFAM; PF00550; pp-binding; 3.
DR PRINTS; PR00154; AMPBINDING.
SQ SEQUENCE 3131 AA; 346891 MW; 0049DB9 CRC32;

Query Match
Best Local Similarity 57.1%; Score 36; DB 3; Length 3131;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RLAIRIXRILLRY 14
        :||:|| |:|:
Db 2637 KLAVRIGRRLIRH 2649

RESULT 8
ID Q44541
AC Q44541; PRELIMINARY; PRT; 265 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ORF 7.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC Azotobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89123097.
RA JACOBSON M.R., BRIGLE K.E., BENNETT L.T., SETTERQUIST R.A.,
RA WILSON M.S., CASH V.L., BEYNON J., NEWTON W.E., DEAN D.R.;
RT "Physical and genetic map of the major nif gene cluster from
RT Azotobacter vinelandii.";
RL J. Bacteriol. 171:1017-1027(1989).
DR EMBL; M20568; AAA64728.1; -.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.
DR PFAM; PF00132; hexapep; 3.
SQ SEQUENCE 265 AA; 28346 MW; 0CE25212 CRC32;

Query Match
Best Local Similarity 55.6%; Score 35; DB 2; Length 265;
Matches 8; Conservative 1; Mismatches 1; Indels 5; Gaps 0;

Qy 1 YRLAIRIXRILLRY 14
       |||| |:| ||
Db 40 YRLANRLWRAAWRY 53

RESULT 9
ID Q57090
AC Q57090; PRELIMINARY; PRT; 402 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TRANSPOSASE.
OS Corynebacterium xerosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
RN [1]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=M82B;
RX   MEDLINE; 96117603.
RA   TAUCH A., KASSING F., KALINOWSKI J., PUHLER A.;
RT   "The Corynebacterium xerosis composite transposon Tn5432 consists of
RT   two identical insertion sequences, designated IS1249, flanking the
RT   erythromycin resistance gene ermCX.";
RL   Plasmid 34:119-131(1995).
DR   EMBL; U21300; AAC95477.1; -.
DR   EMBL; U21300; AAC95474.1; -.
SQ   SEQUENCE   402 AA;   45846 MW;   26339FD8 CRC32;

Query Match             55.6%;  Score 35;  DB 2;  Length 402;
Best Local Similarity   60.0%;  Pred. No. 49;
Matches    6;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YRLAIRIXRI 10
          |||::: ||
Db    210 YRLALKLTRI 219


RESULT 10
Q9XIL2
ID   Q9XIL2         PRELIMINARY;      PRT;     404 AA.
AC   Q9XIL2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE CYCLOPHILIN-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC
DE   5.2.1.8).
GN   F19G14.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
OC   eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA   BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA   CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA   FRASER C.M., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC F19G14 genomic sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
DR   EMBL; AC006438; AAD41985.1; -.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
KW   Isomerase; Rotamase.
SQ   SEQUENCE   404 AA;   45001 MW;   A8A760CD CRC32;

Query Match             55.6%;  Score 35;  DB 10;  Length 404;
Best Local Similarity   50.0%;  Pred. No. 49;
Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 YRLAIRIXRILLRY 14
          |::|:| | |||
Db    228 YKMALRKYRKALRY 241


RESULT 11
050159
ID   050159         PRELIMINARY;      PRT;     469 AA.
AC   050159;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PROTON-TRANSLOCATING ATPASE, BETA SUBUNIT (EC 3.6.1.34).
GN   ATPD.
OS   Streptococcus bovis.


OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JB-1;
RA   UMEMORI J., MIWA T., NAGAMINE T., OGATA K., TAKENAKA A., HINO T.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB09314; BAA23755.1; -.
DR   PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR   PFAM; PF00006; ATP-synt_ab; 1.
DR   PFAM; PF00306; ATP-synt_ab_C; 1.
KW   Hydrolase; Hydrogen ion transport.
SQ   SEQUENCE   469 AA;   51219 MW;   A4170D3F CRC32;

Query Match             55.6%;  Score 35;  DB 2;  Length 469;
Best Local Similarity   42.9%;  Pred. No. 57;
Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YRLAIRIXRILLRY 14
          |:|: :|:||
Db    362 YEVATEVQRVLQRY 375


RESULT 12
049025
ID   049025         PRELIMINARY;      PRT;      44 AA.
AC   049025;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PLASMID GCH7220, COMPLETE SEQUENCE.
OS   Gracilaria chilensis.
OG   Plasmid Gch7220.
OC   Eukaryota; Rhodophyta; Florideophyceae; Gracilariales; Gracilariaceae;
OC   Gracilaria.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GOFF L.J., MOON D.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034719; AAC04733.1; -.
KW   Plasmid.
SQ   SEQUENCE   44 AA;   5487 MW;   6A285329 CRC32;

Query Match             54.0%;  Score 34;  DB 10;  Length 44;
Best Local Similarity   35.7%;  Pred. No. 8.6;
Matches    5;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 YRLAIRIXRILLRY 14
          ||: ::: :|| |
Db     23 YRMCVKVKKILYFY 36


RESULT 13
Q9YC92
ID   Q9YC92         PRELIMINARY;      PRT;     157 AA.
AC   Q9YC92;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   157AA LONG HYPOTHETICAL PROTEIN.
GN   APE1362.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
```

RA   YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000061; BAA80356.1; -.
SQ   SEQUENCE   157 AA; 17942 MW; C1F4AB62 CRC32;

Query Match   54.0%; Score 34; DB 1; Length 157;
Best Local Similarity   58.3%; Pred. No. 30;
Matches   7; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   3 LAIRIXRILLRY 14
            || :| |||::|
Db   35 LAAKIARILTKY 46

RESULT 14
P94457
ID   P94457   PRELIMINARY;   PRT;   170 AA.
AC   P94457;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ORF4 GENE (FRAGMENT).
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 8224;
RX   MEDLINE; 94113715.
RA   SAKANYAN V., DESMAREZ L., LEGRAIN C., CHARLIER D., METT T.,
RA   KOCHIKYAN A., SAVCHENKO A., BOYEN A., FALMAGNE P., PIRARD A.,
RA   GLANSDORFF N.;
RT   "Gene cloning, sequence analysis, purification, and characterization
RT   of a thermostable aminoacylase from Bacillus stearothermophilus.";
RL   Appl. Environ. Microbiol. 59:3878-3888(1993).
DR   EMBL; Y08751; CAA69998.1; -.
FT   NON_TER   1   1
SQ   SEQUENCE   170 AA; 18433 MW; 60A9C69B CRC32;

Query Match   54.0%; Score 34; DB 2; Length 170;
Best Local Similarity   50.0%; Pred. No. 33;
Matches   7; Conservative   2; Mismatches   2; Indels   5; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
            ||| :: ||||
Db   32 YRLEMKTTEKLLRY 45

RESULT 15
Q9YFD0
ID   Q9YFD0   PRELIMINARY;   PRT;   276 AA.
AC   Q9YFD0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   276AA LONG HYPOTHETICAL CITRATE LYASE BETA CHAIN.
GN   APE0311.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA   YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000059; BAA79266.1; -.
KW   Lyase.
SQ   SEQUENCE   276 AA; 30725 MW; 4B749B21 CRC32;

Query Match   54.0%; Score 34; DB 1; Length 276;
Best Local Similarity   50.0%; Pred. No. 53;
Matches   7; Conservative   2; Mismatches   2; Indels   5; Gaps   0;

Qy   1 YRLAIRIXRILLRY 14
            |||: :| ||
Db   259 YRLALNLLRRASRY 272

Search completed: February 8, 2000, 13:17:49
Job time: 32498 sec

THIS PAGE BLANK (USPTO)

OM of: US-08-653-294-20 to: GenEmbl:*   out_format : pfs

Date: Feb 8, 2000   4:44 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15779/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=-4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-20
Query length: 14
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 11370.480000

score_list:

sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_ro:AC002121	-	43.00	101.77	1.2e+03	84056	! AC002121 Genomic sequence fro
gb_ba2:APU89957	+	42.00	120.19	113.36	6715	! U89957 Actinobacillus pleuropn
gb_pr2:AP000657	+	42.00	102.43	1.1e+03	50660	! AP000657 Homo sapiens genomic
gb_htg6:AC013287	+	42.00	91.66	4.4e+03	172450	! AC013287 Homo sapiens chromoso
gb_pl1:SCYPL167C	+	41.00	119.30	127.11	4828	! 273523 S.cerevisiae chromosome
gb_pr1:HUMOPIODRE	-	41.00	119.28	127.43	4839	! M84605 Human putative opioid r
gb_pl1:YSCREV3	+	41.00	118.90	133.89	5056	! M29683 S.cerevisiae DNA polyme
gb_pl1:SCLACHXVI	+	41.00	97.79	2.0e+03	55786	! X96770 S.cerevisiae chromosom
gb_htg5:AC013211	-	41.00	117.79	154.41	3727	! AC013211 Drosophila melanogast
gb_inl:CEE02A10	+	40.00	101.45	1.3e+03	23889	! Z81053 Caenorhabditis elegans
gb_inl:CEM01F1	-	40.00	97.76	2.0e+03	36355	! Z46381 Caenorhabditis elegans
gb_inl:CET21C9	+	40.00	97.48	2.1e+03	37545	! Z73098 Caenorhabditis elegans
gb_pl1:AB028622	+	40.00	90.57	5.1e+03	82348	! AB028622 Arabidopsis thaliana
gb_pl2:ATAC007070	+	40.00	90.19	5.3e+03	86017	! AC007070 Arabidopsis thaliana
gb_htg3:AC004548	+	40.00	88.89	6.3e+03	99657	! AC004548 Homo sapiens chromos
gb_htg1:CEY24E3	+	40.00	86.11	9.0e+03	136764	! AL008873 Caenorhabditis eleg
gb_pr2:AC002456	+	40.00	84.93	1.0e+04	156399	! AC002456 Human BAC clone RGC
gb_htg2:AL133355	-	40.00	83.17	1.3e+04	191137	! AL133355 Homo sapiens clone
gb_htg6:AC012517	-	40.00	81.39	1.6e+04	233904	! AC012517 Drosophila melanogas
gb_pl2:ABU54809	+	39.00	128.15	40.88	745	! U54809 Armillaria borealis rDNA
gb_pr3:AF056334	-	39.00	112.81	292.39	4265	! AF056334 Homo sapiens cancer/t
gb_bal:MVOMCR	-	39.00	110.71	382.48	5412	! M16893 Methanococcus vannielii
gb_pr3:AF064589	+	39.00	109.14	467.87	6471	! AF064589 Homo sapiens melanoma
gb_bal:MVMCR1	-	39.00	109.14	467.95	6472	! X07793 Methanococcus voltae mc
gb_bal:BSPOLKET	+	39.00	107.57	572.14	7735	! 235133 B.subtilis 168 pks gene
gb_htg7:AC018125	-	39.00	106.13	688.57	9116	! AC018125 Drosophila melanogast
gb_htg7:AC017410	-	39.00	95.72	3.0e+03	29776	! AC017410 Drosophila melanogas
gb_htg2:AC006614	+	39.00	94.61	3.6e+03	33803	! AC006614 Caenorhabditis elega
gb_htg5:AC015441	+	39.00	94.46	3.1e+03	34357	! AC015441 Drosophila melanogas
gb_bal:MTCY373	-	39.00	94.17	3.2e+03	35516	! 273419 Mycobacterium tubercul
gb_bal:MLU15186	+	39.00	93.99	3.3e+03	36241	! U15186 Mycobacterium leprae c
gb_in2:AC005447	+	39.00	87.27	7.7e+03	77907	! AC005447 Drosophila melanogas
gb_pl2:ATAC007187	+	39.00	85.21	1.0e+04	98432	! AC007187 Arabidopsis thaliana
gb_pl2:ATF4IIO	+	39.00	84.08	1.2e+04	111876	! AL035525 Arabidopsis thalian
gb_pr2:CNS01DRM	+	39.00	83.60	1.2e+04	118261	! AL117667 Human chromosome 14
gb_htg2:AC008259	+	39.00	83.49	1.3e+04	119692	! AL008259 Drosophila melanoga
gb_pr3:HS326L12	-	39.00	83.44	1.3e+04	120423	! AL023279 Homo sapiens DNA se
gb_pr2:HS232G24	+	39.00	83.14	1.3e+04	124497	! AL022152 Homo sapiens DNA se
gb_htg3:AC009393	-	39.00	82.92	1.4e+04	124727	! AC009393 Drosophila melanoga
gb_pr3:HS769D20	+	39.00	82.50	1.4e+04	133889	! AL031643 Human DNA sequence
gb_pr4:AC007262	+	39.00	80.92	1.7e+04	160359	! AC007262 Homo sapiens chromc
gb_htg5:AC011614	+	39.00	80.30	1.9e+04	172043	! AC011614 Drosophila melanoga

gb_pr2:CNS0000E   -   39.00   79.54   2.1e+04 187564   ! AL049835 Human chromosome
gb_bal:BSUB0010   +   39.00   77.60   2.6e+04 233780   ! Z99113 Bacillus subtilis
gb_pl1:AB017564   +   38.00 121.92   90.89   983   ! AB017564 Arabidopsis thalian

seq_name: gb_ro:AC002121

seq_documentation_block:
LOCUS   AC002121   84056 bp   DNA   ROD   10-JUL-1997
DEFINITION Genomic sequence from Mouse 11, complete sequence.
ACCESSION   AC002121
VERSION   AC002121.1 GI:2133880
KEYWORDS   HTG.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 84056)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H. and Lander,E.S.
  TITLE   Genomic sequence from Mouse 11
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 84056)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 84056)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 84056)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUL-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 29, 1997 this sequence version replaced gi:2098549.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES   Location/Qualifiers
     source   1..84056
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="S157"
                     /clone_lib="unknown"
                     /chromosome="19"
     repeat_region   complement(4..117)

```
repeat_region                 /rpt_family="B3"
repeat_region  complement(128..317)   /rpt_family="B2"
repeat_region  complement(321..438)   /rpt_family="PB1D7"
repeat_region  complement(714..841)   /rpt_family="B1-F"
repeat_region  905..1051              /rpt_family="B1_MM"
repeat_region  complement(1051..1109) /rpt_family="4.5SRNA"
repeat_region  1536..1609             /rpt_family="BC1_MM"
repeat_region  1714..1747             /rpt_family="AT_rich"
repeat_region  complement(1817..1920) /rpt_family="B1_MM"
repeat_region  1946..2038             /rpt_family="B1_MM"
repeat_region  2070..2179             /rpt_family="PB1D10"
repeat_region  2648..2748             /rpt_family="PB1D10"
repeat_region  5761..5953             /rpt_family="MER58A"
repeat_region  6086..6138             /rpt_family="(CA)n"
repeat_region  6432..6507             /rpt_family="PB1D10"
repeat_region  6643..6766             /rpt_family="B1-F"
repeat_region  6809..6841             /rpt_family="ORR1B-INT"
repeat_region  6842..6988             /rpt_family="B1_MM"
repeat_region  7020..7047             /rpt_family="(GA)n"
repeat_region  complement(7679..7715) /rpt_family="(CCA)n"
repeat_region  complement(7716..7862) /rpt_family="B1-F"
repeat_region  complement(8122..8228) /rpt_family="PB1D7"
repeat_region  complement(8246..8318) /rpt_family="ID4"
repeat_region  complement(8654..8679) /rpt_family="AT_rich"
repeat_region  10234..10300           /rpt_family="PB1D7"
repeat_region  10323..10409           /rpt_family="ID_RN"
repeat_region  10334..10430           /rpt_family="BC1_MM"
repeat_region  complement(11042..11114) /rpt_family="ID2"
repeat_region  11146..11271           /rpt_family="B1_MM"
repeat_region  complement(11318..11391) /rpt_family="ID4"
repeat_region  11520..11636           /rpt_family="(CATA)n"
repeat_region  complement(11657..11708) /rpt_family="AT_rich"
repeat_region  complement(11768..11934) /rpt_family="RSINE1"
repeat_region  12720..12818           /rpt_family="B1-F"
repeat_region  12840..12994           /rpt_family="RSINE1"
repeat_region  12991..13151           /rpt_family="RSINE2"
repeat_region  13021..13139           /rpt_family="B1_MM"

repeat_region  13180..13355           /rpt_family="B3"
repeat_region  13395..13436           /rpt_family="RSINE2"
repeat_region  13410..13443           /rpt_family="PB1D9"
repeat_region  13459..13529           /rpt_family="B4A"
repeat_region  complement(13824..13903) /rpt_family="B1-F"
repeat_region  complement(15464..15608) /rpt_family="ID5"
repeat_region  complement(15618..15808) /rpt_family="B1_MM"
repeat_region  15936..16159           /rpt_family="B2"
repeat_region  16169..16354           /rpt_family="B3"
repeat_region  18144..18206           /rpt_family="RSINE1"
repeat_region  18267..18368           /rpt_family="ID4"
repeat_region  18496..18587           /rpt_family="B1_MM"
repeat_region  18580..18706           /rpt_family="B4A"
repeat_region  18612..18806           /rpt_family="ORR1B-INT"
repeat_region  18612..18728           /rpt_family="B4"
repeat_region  19217..19365           /rpt_family="B1_MM"
repeat_region  19481..19558           /rpt_family="B1_MM"
repeat_region  20582..20677           /rpt_family="PB1D7"
repeat_region  complement(20723..20750) /rpt_family="(GAAAA)n"
repeat_region  21179..21292           /rpt_family="POLY_A"
repeat_region  21328..21475           /rpt_family="PB1D10"
repeat_region  21493..21626           /rpt_family="B1_MM"
repeat_region  complement(21831..21879) /rpt_family="B1-F"
repeat_region  complement(22418..22439) /rpt_family="AT_rich"
repeat_region  22532..22635           /rpt_family="POLY_A"
repeat_region  complement(25305..25400) /rpt_family="MIR"
repeat_region  complement(26260..26457) /rpt_family="B1-F"
repeat_region  complement(26665..26707) /rpt_family="B2"
repeat_region  26793..26908           /rpt_family="(CATA)n"
repeat_region  complement(26911..26959) /rpt_family="B1_MM"
repeat_region  complement(26960..27022) /rpt_family="(GAAAA)n"
```

```
alignment_scores:
            Quality:  43.00             Length:  14
              Ratio:  3.583               Gaps:   0
 Percent Similarity: 85.714   Percent Identity: 64.286

alignment_block:
us-08-653-294-20 x AC002121/rev

Align seg 1/1 to reverse of: AC002121  from: 1  to: 84056

  1 TyrLeuAlaIleArgIle**ArgIleLeuArgTyr 14
```

```
             |||||  |||::  |||||||||::::||||||||||||||:::
      44056 TATAGATCAGCAGTGACAATTGCCAGGGTCCTGCTCAGGCAC 44015

seq_name: gb_ba2:APU89957

seq_documentation_block:
LOCUS       APU89957     6715 bp    DNA             BCT       03-FEB-1998
DEFINITION  Actinobacillus pleuropneumoniae urease operon (ureABCXEFGD) genes,
            complete cds.
ACCESSION   U89957
VERSION     U89957.1  GI:2828540
KEYWORDS    .
SOURCE      Actinobacillus pleuropneumoniae.
ORGANISM    Actinobacillus pleuropneumoniae
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Actinobacillus.
REFERENCE   1  (bases 1 to 6715)
AUTHORS     Bosse,J.T. and MacInnes,J.I.
TITLE       Urease genes of Actinobacillus pleuropneumoniae
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 6715)
AUTHORS     Bosse,J.T. and MacInnes,J.I.
TITLE       Direct Submission
JOURNAL     Submitted (18-FEB-1997) Pathobiology, University of Guelph, Guelph,
            Ont N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1..6715
                     /organism="Actinobacillus pleuropneumoniae"
                     /strain="CM5"
                     /db_xref="taxon:715"
     gene            184..486
                     /gene="ureA"
     CDS             184..486
                     /gene="ureA"
                     /codon_start=1
                     /transl_table=11
                     /product="UreA"
                     /protein_id="AAC00058.1"
                     /db_xref="GI:2828541"
                     /translation="MALITPRSQYVATYGPTVGDKVRLGDTDLWATIEQDFLTKGDEC
                     EAARDGMSVAEVMQYGATLLTVDDVMGGIAEMVHEVQIEATFPDGTKLVTVHNPIR"
     gene            676..981
                     /gene="ureB"
     CDS             676..981
                     /gene="ureB"
                     /codon_start=1
                     /transl_table=11
                     /product="UreB"
                     /protein_id="AAC00059.1"
                     /db_xref="GI:2828542"
                     /translation="MIPGEYQLADGDVQANVGRKTVKLEVVNSGDRPIQVGSHYHFFE
                     TNHALKFDRLQARGMRLNVPSGNAVRFEPGEAKEVELVEFGGNKVIYGFHNEIDGKL"
     gene            993..2711
                     /gene="ureC"
     CDS             993..2711
                     /gene="ureC"
                     /codon_start=1
                     /transl_table=11
                     /product="UreC"
                     /protein_id="AAC00060.1"
                     /db_xref="GI:2828543"
                     /translation="MALTIPRSQYVATYGPTVGDKVRLGDTDLWATIEQDFLTKGDEC
                     KFGGGKSVRDGMAQSSTATRDNPNVLDFALTNVMIIDAKLGIIKADIGIRDGRIVGIG
                     QAGNPDTMDNVTPNMIIGASTEVHNGAHLIATAGGIDTHIHWICPQQAQHAIENGITT
                     MIGGGSGPADGTHATTCTPGKFNIERMFQACEALPVNSGITAGLGLEPLKEQVVA
                     GALGLKIHEDWGATPAVIDAALKVADEMDVQVAIHTDTLNESGFLEDTMKAINGRVIH
                     TFHTEGAGGGHAPDIIKAAMYPNVLPASTNPNRPFTVNTIDEHLDMLMVCHHLDKRVP
                     EDVAFADSRIRPETIAAEDILHDMGVFSIMSSDSQAMGRVGEVVTRTWQTADKMKAQR
                     GALGDEGNDNFRIKRYIAKYTINPAIAHGISQYVGSLEVGKLADIVLMKPQFFGVKPE
                     FVMKKGFISFAKMGDPNASIPTTPQVFYRPMFGANAKANTESAVYFVSQASVDANIKA
                     QYGIQKETLAVKGCRDVCKKDLVHNNATPEITVDTERYEVRVDGEHITCEPATKVPLA
                     QRYLF"
     gene            2835..3452
                     /gene="ureX"
```

```
     CDS             2835..3452
                     /gene="ureX"
                     /codon_start=1
                     /transl_table=11
                     /product="UreX"
                     /protein_id="AAC00061.1"
                     /db_xref="GI:2828544"
                     /translation="MKLWYSTTSPFVRKVLVTLKHQQLEAKTELLKISSSFDPNSPHN
                     QDNPLGRVPALQRNCGNWLFGSQLICEYLDQKGDQPKLFPESGKFRWAALALHNLADG
                     ILENTVPIMAERLLRPENEWTSRQEQLMERNIRSFPQLEKAIEPFGTELNIGIITAV
                     CLIDWLQFRAEKLGYDLAKNHPNLTAWAEEMNGKYAVLAETKPHV"
     gene            3568..4146
                     /gene="ureE"
     CDS             3568..4146
                     /gene="ureE"
                     /codon_start=1
                     /transl_table=11
                     /product="UreE"
                     /protein_id="AAC00062.1"
                     /db_xref="GI:2828545"
                     /translation="MKNTIMQILNPILPVMEDILGNLAELKASGKITNQQIDTVELQW
                     YESERNILRKTSKSGREVAFRLLKEGQRLKHDDVVFVSELVIAEILPSEVIVLSPK
                     TLPEMARACYEIGNKHSPLFLDGDEVTLPYDKPMFEWLQAAGFGPQKAERRLSQALRA
                     NSAQHGHSHGHSHSHDHGYHHHGDGNWHKH"
     gene            4158..4841
                     /gene="ureF"
     CDS             4158..4841
                     /gene="ureF"
                     /codon_start=1
                     /transl_table=11
                     /product="UreF"
                     /protein_id="AAC00063.1"
                     /db_xref="GI:2828546"
                     /translation="MGQLGALLHLVDPTLPIGGFNHSNGLETFVQQGKVNSRASLEEY
                     VQTQLMQNWIYNDGAYLSLAFDAMANHDLDRLLALDQELAASKIARESREGSYKLGVR
                     LLKIFIRYENHPLLSEFQQAVSEKRCQGYFPIVFAMVAQAMNLDKAETLYAFYYNAAV
                     GVVTNGVKLVPLSQMDGQDILFALRTPLAQAVENSLNPDLDWLGAATLASDIRSMQHE
                     QLYTRLYMS"
     gene            4862..5497
                     /gene="ureG"
     CDS             4862..5497
                     /gene="ureG"
                     /codon_start=1
                     /transl_table=11
                     /product="UreG"
                     /protein_id="AAC00064.1"
                     /db_xref="GI:2828547"
                     /translation="MRKYIKIGVAGPVGAGKTALIERLTREIASKYSVAVITNDIYTQ
                     EDAEFLTKNSLLPPERIMGVETGGCPHTAIREDASMNLEAVDEMVARFPEVELIFES
                     GGDNLSATFSPDLADVTIFVIDVAQGEKIPRKGPGISRSDLLVINKTDLAPFVGADL
                     SVMERDARRMRNGQPFIFTNLMKNENLDGVIGWIEKYALLKNIEDPASLVR"
     gene            5666..6478
                     /gene="ureD"
     CDS             5666..6478
                     /gene="ureD"
                     /codon_start=1
                     /transl_table=11
                     /product="UreD"
                     /protein_id="AAC00065.1"
                     /db_xref="GI:2828548"
                     /translation="MQSKLLSTKLTSQGKTQLDQYFVSPPFKVMTLPAYDDAWQNGL
                     NAMQMSSPGLLASDLLDIEISLADDTALSLNTQAFTRVQSMNEGDYATQKTCIKLGK
                     NSRLFYLPHPLVLHKDSSFKQTTEIEMSEQSELIYGEIVAIGRVLNGERFAFRHFASY
                     LRISYQNRPIADRIQWLPAKMALTSLSQMEDFSHQGSLTYVNLAKNAVEIKAMVSEL
                     QALAAEQKNMLIGVSQLNEGGLMVRVLAHRADIIQHLFERIGQVLKAQSNIV"
BASE COUNT     1972 a   1408 c   1642 g   1693 t
ORIGIN

alignment_scores:
        Quality:    42.00       Length:      14
        Ratio:       3.231       Gaps:         0
```

```
Percent Similarity: 92.857     Percent Identity: 42.857

alignment_block:
US-08-653-294-20 x APU89957

Align seg 1/1 to: APU89957 from: 1 to: 6715

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
      ||:::||:::::|||::   :::|||::|||::::::|||||
 4446 TATAAATTAGGTGTACGATTGCTGAAGAGATTTTTATTCGTTAT 4487

seq_name: gb_pr2:AP000657

seq_documentation_block:
LOCUS       AP000657    50660 bp    DNA          PRI    20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.1-q21.2, clone:B662K22,
            LL56-APP region, complete sequence.
ACCESSION   AP000657
VERSION     AP000657.1  GI:6252989
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:B662K22.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 50660)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 50,660 genomic DNA of 21q21.1-q21.2
JOURNAL     Published Only in DataBase (1999) In press
REFERENCE   2 (bases 1 to 50660)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
FEATURES    Location/Qualifiers
     source  1..50660
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="21"
             /clone="B662K22"
             /map"21q21.1-q21.2"
BASE COUNT  16359 a   9174 c   8753 g   16374 t
ORIGIN

alignment_scores:
    Quality:     42.00        Length:        12
    Ratio:       3.500        Gaps:           0
Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
US-08-653-294-20 x AP000657

Align seg 1/1 to: AP000657 from: 1 to: 50660

    3 LeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
      |||||:::||:::::|||::|||||||||||||||||
11470 CTTGCTGTGAAGGTCTCCAAAATTCTTTTGAGGTAT 11505

seq_name: gb_htg6:AC013287

seq_documentation_block:
LOCUS       AC013287   172450 bp    DNA          HTG        08-DEC-1999
DEFINITION  Homo sapiens clone RP11-76F14, *** SEQUENCING IN PROGRESS ***, 28
            unordered pieces.
ACCESSION   AC013287
VERSION     AC013287.3  GI:6539165
KEYWORDS    HTG; HTGS_PHASE1.
```

```
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 172450)
AUTHORS     Smith,D.R.
TITLE       Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 172450)
AUTHORS     Smith,D.R.
TITLE       Direct Submission
JOURNAL     Submitted (06-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Dec 8, 1999 this sequence version replaced gi:6456128.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 28 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1107: contig of 1107 bp in length
          *     1108     2200: gap of unknown length
          *     2201     4161: contig of 1961 bp in length
          *     4162     5428: contig of 1267 bp in length
          *     5429     6691: gap of unknown length
          *     6692     8694: contig of 2003 bp in length
          *     8695    10885: contig of 2191 bp in length
          *    10886    13908: contig of 3023 bp in length
          *    13909    16572: contig of 2664 bp in length
          *    16573    19009: contig of 2437 bp in length
          *    19010    21781: contig of 2772 bp in length
          *    21782    25274: contig of 3493 bp in length
          *    25275    27762: contig of 2488 bp in length
          *    27763    30099: contig of 2337 bp in length
          *    30100    34346: contig of 4247 bp in length
          *    34347    39593: contig of 5247 bp in length
          *    39594    44752: contig of 5159 bp in length
          *    44753    50656: contig of 5904 bp in length
          *    50657    55411: contig of 4755 bp in length
          *    55412    60112: contig of 4701 bp in length
          *    60113    68267: contig of 8155 bp in length
          *    68268    77608: contig of 9341 bp in length
          *    77609    88341: contig of 10733 bp in length
          *    88342   101401: contig of 13060 bp in length
          *   101402   113276: contig of 11875 bp in length
          *
```

```
     *       113277   128180: contig of 14904 bp in length
     *                        gap of unknown length
     *       128181   145821: contig of 17641 bp in length
     *                        gap of unknown length
     *       145822   172450: contig of 26629 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-76F14"
BASE COUNT  55457 a   31143 c   31194 g   54614 t   42 others
ORIGIN

alignment_scores:
        Quality:    42.00              Length:      14
        Ratio:      3.231              Gaps:         0
Percent Similarity: 92.857   Percent Identity: 50.000

alignment_block:
US-08-653-294-20 x AC013287

Align seg 1/1  to: AC013287  from: 1  to: 172450

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
      |||:|||::|||::||:|||    |||||||||||||||||
15208 TATAACATATCATCAATTTCACAGAATTCTAATGAGATAT 15249

seq_name: gb_pl1:SCYPL167C

seq_documentation_block:
LOCUS       SCYPL167C     4828 bp       DNA              PLN      11-AUG-1997
DEFINITION  S.cerevisiae chromosome XVI reading frame ORF YPL167c.
ACCESSION   Z73523 U00094
VERSION     Z73523.1  GI:1370352
KEYWORDS    .
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1 (bases 1 to 4828)
AUTHORS     Purnelle,B., Coster,F. and Goffeau,A.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 4828)
AUTHORS     MIPS.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XVI sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .4828
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            complement(84. .4598)
                     /gene="REV3"
     CDS             complement(84. .4598)
                     /gene="REV3"
                     /note="ORF YPL167c"
                     /codon_start=1
                     /db_xref="SGD:S0006088"
                     /db_xref="SGD:S0006088"
                     /protein_id="CAA97873.1"
                     /db_xref="GI:1370353"
                     /db_xref="SWISS-PROT:P14284"
                     /translation="MSRESNDTIQSDFVRSSSKSDYFRIQLNNQDYYMSKPTFLDPSH
                     GESLPLNQFSQVPNIRVFGALPTGHQVLCHVHGILPYMFIKYDGQITDTSLRHQRCA
                     QVHKTLEVKIRASFKRKKDKHDLAGDKLGNLNFVADVSVVKGIPFYGYHVGWNLFYK
                     ISLLNPSCLSRISELIRDGKIFGKKFEIYESHIPYLLQWTADFNLFGCSWINVDRCYF
                     RSPVLNSILEIDIKLTINDDLQLLDRFCDFKCNVLSRRDFPRVGNGLIEIDILPQFIK
                     NREKLQHRDIHHDFLEKLGDISDIPVKPYVSSARDMINELTMQREELSLKEYKEPPET
                     KRHVSGHQWQSSGEFAFYKAQHKTSTFDGQIPNFENFIDKNQKFSAINTPYEALPQ
```

```
LWPRLPQIEINNSMQDKKNDQVNASTEYEICGVDNENEGVKGSNIKSRSYSWLPE
SIASPKDSTILLDHQTKYHNTINFSMDCAMTONMASKRKLRSSVSANKTSLLSRKRKK
VMAAGLRYGKRAFVYGEPPFGYQDILNKLEDEGFPKIDYKDPFFSNPVDLENKPYAYA
GKRFEISSTHVSTRIPVQFGGETVSVXNKPTFDMFSSWKYALRPPTYDAVQKWYNKVP
SMGNKKTESQISMHTPHSKFLYKFASDVSGKQKRKKSSVHDSLTHLTLEIHANTRSDK
IPPDAIDEVSMIIWCLEETFPLDLDIAYEGIMIVHKASEDSTFPTKIQHCINEIPVM
FYESEFMFEALTDLVLLLDPDILSGFEIHNFSWGYIIERCQKIHQFDIVRELARVKC
QIKTKLSDTWGYAHSSGIMITGRHMINTWRALRSDVNLTQYTIESAAFNIÐHKRLPHF
SFESLTNMWNAKKSTTELKTVLNYWLSRAQINIQOLLRKQDYIARNIEQARLIGIDFHS
VYYRGSQFKVESFLIRICKSESFILLSPGKKDVRKQKALECVPLVMEPESAFYKSPLI
VLDFQSLYPSIMIGYNYCYSTMIGRVREINLTENNLGVSKFSLPRNILALLKNDVTIA
PNGVVYAKTSVRKSTLSKMLTDILDVRVMIKTMNEIGDDNTTEKRLNNKQLALKLL
ANVTYGYTSASFSGRMPCSDLADSIVQTGRETLEKAIDIIEKDETWNAKVVYGDTDSL
FVYLPGKTAIEAFSIGHAMAERVTQNNPKPIFLKFEKVYHPSILISKKRYVGFSYESP
SQTLPIFDAKGIETVRRDGIPAQOKIIEKCIRLLFQFKDLSIKKYÐNEFFKIQIGK
VSAQDFCFAKEVKLGAYKSEKTAPAGAVVKRRINEDHRAEPQYKERIPYLVVKGKQG
QLLRERCVSPEEFLEGENELDSEYINKILIPPLDRLFNLIGINVGNWAQEIVKSKR
ASTTTTKVENITRVGTSATCCNCGEELTKICSLQLCDDCLEKRSTTLSFLIKKLKRQ
KEYQTLKTVCRTCSYRYISDAGIENDHIASKCNSYDCPVFYSRVKAERYLRDNQSVQR
EEALISLNDW

BASE COUNT  1381 a   927 c   818 g   1702 t
ORIGIN

alignment_scores:
        Quality:    41.00              Length:      14
        Ratio:      3.727              Gaps:         0
Percent Similarity: 78.571   Percent Identity: 57.143

alignment_block:
US-08-653-294-20 x SCYPL167C/rev

Align seg 1/1  to reverse of: SCYPL167C  from: 1  to: 4828

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
      ||||||  |||::||::|||:    ::|||||||||||  |||
 3256 TACCGGAAAGCATTGCGTCCCGAAAGATTCTACTATACTAT 3215

seq_name: gb_prl:HOMOPIODRE

seq_documentation_block:
LOCUS       HOMOPIODRE     4839 bp       mRNA            PRI       27-APR-1993
DEFINITION  Human putative opioid receptor mRNA, complete cds.
ACCESSION   M84605
VERSION     M84605.1  GI:189391
KEYWORDS    opioid receptor.
SOURCE      Homo sapiens female placenta cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4839)
AUTHORS     Goldstein,A., Xie,G.-X. and Miyajima,A.
TITLE       Expression cloning of cDNA encoding a seven-helix receptor from
            human placenta with affinity for opioid ligands
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 89, 4124-4128 (1992)
MEDLINE     92237319
COMMENT     Transmembrane helices (numbers are AA residues):
              I     60. . .82
             II     93. . .114
            III    135. . .156
             IV    177. . .197
              V    221. . .245
             VI    275. . .296
            VII    310. . .334
            serine-rich cytoplasmic tail:     355. . .440
            Possible glycosylation sites:     7, 24.
FEATURES             Location/Qualifiers
     source          1. .4839
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /tissue_type="placenta"
     CDS             94. .1416
```

```
              /codon_start=1
              /product="putative opioid receptor"
              /protein_id="AAA36395.1"
              /db_xref="GI:189392"
              /translation="MASPAGNLSAWPGWGWPPPAALRNLTSSPAPTASPSPAPSWTPS
              PRPGPAHPFLQPPWAVAVAWSLAVGAVVAVAVLGNLVVIWIVLAHKRMRTVTNSFLVNL
              AFADAMALNALVNFTYALHGEWYFGANYCRFQNFFPITAVFASIYSMTAIAVDRYM
              AIIDPLKPRLSATATRIVIGSIWILAFLLAFPQCLYSKIKVMPGRTLCVVQWPEGSRQ
              HFTYHMIVIVLVCFPLLIMGITYTIVIGITLWGGEIPGDTCDRYQEQLKAKRKVVKMM
              IIVVTFAICWLPIHIYFILTAIYQQLNRWKYIQQVYLASFWLAMSSTMYNPIIYCCL
              NKRFRAGFKRAFRWCPFLYSSYDELELAKATRLHPMRQSSLYTVTRMEMSVVFDSND
              GDSARSSHQKRGTTRDVGSNVCSRRNSKSTSTTASFVSSSHMSVEEGS"
BASE COUNT     1449 a   1099 c    965 g   1326 t
ORIGIN

alignment_scores:
        Quality:    41.00              Length:     13
        Ratio:       3.417             Gaps:        0
Percent Similarity: 92.308     Percent Identity: 76.923

alignment_block:
US-08-653-294-20 x HUMOPIODRE

Align seg 1/1  to: HUMOPIODRE  from: 1  to: 4839

       2 ArgLeuAlaIleArgIle***ArgIleLeuArgTyr 14
         ||||||  |||::||||||:::|||||||||||||||
    1530 AGATTGCACATAAATAACAAAAATACTACTAAGADATAT 1568

seq_name: gb_pl1:YSCREV3

seq_documentation_block:
LOCUS       YSCREV3       5056 bp    DNA           PLN       27-APR-1993
DEFINITION  S.cerevisiae DNA polymerase (rev3) gene, complete cds.
ACCESSION   M29683
VERSION     M29683.1  GI:172386
KEYWORDS    DNA polymerase.
SOURCE      S.cerevisiae DNA.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5056)
AUTHORS     Morrison,A., Christensen,R.B., Alley,J., Beck,A.K., Bernstine,E.G.,
            Lemontt,J.F. and Lawrence,C.W.
TITLE       REV3, a Saccharomyces cerevisiae gene whose function is required
            for induced mutagenesis, is predicted to encode a nonessential DNA
            polymerase
JOURNAL     J. Bacteriol. 171, 5659-5667 (1989)
MEDLINE     90008808
FEATURES    Location/Qualifiers
     source  1..5056
             /organism="Saccharomyces cerevisiae"
             /db_xref="taxon:4932"
     CDS     305..4819
             /note="DNA polymerase (pot.); putative"
             /codon_start=1
             /protein_id="AAA34968.1"
             /db_xref="GI:172387"
             /translation="MSRESNDTIQSDTVRSSSKSDYFRIQLNNQDYYMSKPTFLDPSH
             GESLPLNQFSQVPNIRVFGALPTGHQVLCHVHGILPYMFIKYDGQITDTSTLRHQRCA
             QVHKTLENVLKRASFEKRKKDDKHDLAGDKIGNLNFVADVSVVKGIPFYGIHVGWNLFYK
             ISLLNPSCLSRISELIRDGKIFGKKFEIYESHIPYLLQWTADFNLFGCSWINVDRCYF
             RSPVLNSILDIDIKLTINDDLQLLDRFCDFKCNVLSRRDFPRVGNGLIEIDILPQFIK
             NREKLQHRDIHHDFLEKLGLDIGIPCKPVVSSARDMINELTMQREELSLKEYKPPET
             KRHVSGHQWQSSGEFEAFYKKAQHKTSFDGQIPNFENFIDKNQKFSAINTPYEALPQ
             LWPRLPQIEINNNSMQDKKDDQVNASFTEYEICGVDNENEGVKGSNIKSRSYSWLPE
             SIASPKDSTILLDHGTKYHTINFSMDCAMTQNMASKRLRSSVSANKTSLLSKKRKK
             VMAAGLRYGKRAFVYGEPPFGYODILNKLEDEGFPKIDYKDPFFSNPVDLENKPYAYA
             GKRFEISSTHVSTRIPVQFGGETVSYVINKPTFDMFSSWKXALKPPTVDAVQKWYNKVP
             SMGNKKTESOISMHTPHSKFLYKFASDVSGKQKRKKSVHDSLTHLTLEIHANTRSDK
             IPDPAIDEVSMIIWCLEEETFPLDDIAYEGIMVHKASEDSTFPTKIQHCINEIPVM
             FYESEFEMFEALTDIVLLLDPDILSGFEIHNFSWGYIIERCQKIHQFDIVRELARVKC
```

```
              QIKIKLSDTWGYAHSSGIMTGRHMINTWRALRSDVNLTQYTIESAAFNILHKRLPHF
              SFESLTNMWNAKKSTELKTIKVLNVLWLSRAQINIQLLRKQDYIARNIEQARLIGIDFHS
              VYYRGSQFKVESFLRICKSESFLLSPGKKDVRKQKALECVPLVMEPESAFYKSPLI
              VLDFQSLYFSIMIGYNYCISTMIGRVREINTENNLGVSKFSLPRNILALLKNDVTIA
              PNGVVYAKTSVRKSTLSKMLTDILDVRVMIKKTMNEIGDDNTTLKRLLNNKQLALKLL
              ANVYGYTSASFSGRMPCSDLADSIVQTGRETLEKAIDIIEKDETWNAKVVYGDTDSL
              FVYLPGKTAIEAFSIGHAMAERVTQNNPKRPIFLFEKVYHPSILISKKRYVGFSYESP
              SQTLPIFDAKGIETVRRDGIPAQOKIIEKCIRLLFQTKDLSKIKYLQNEFFKIQGK
              VSAADPCFAKEVKIGAYKSEKTAPAGAVVVKRRINEDHRAEPQYKERIPYLVVKGKQG
              QLLRERCVSPEEFLEGENLELDSEYYINKLIIPPLDRLFNLIGINVGNWAQEIVKSKR
              ASTTTTKVENITRVGTSATCCNCGELTKICSLQLCDDCLEKRSTTLSFLIKKLRRQ
              KEYOTLKTVCRTCSYRYTSDAGIENDHIASKCNSYDCPVFYSRVKAERYLRDNQSVQR
              EEALISLNDW"
BASE COUNT     1770 a    856 c    972 g   1458 t
ORIGIN         hromosome XVI.

alignment_scores:
        Quality:    41.00              Length:     14
        Ratio:       3.727             Gaps:        0
Percent Similarity: 78.571     Percent Identity: 57.143

alignment_block:
US-08-653-294-20 x YSCREV3

Align seg 1/1  to: YSCREV3  from: 1  to: 5056

       1 TyrArgAlaIleArgIle***ArgIleLeuLeuArgTyr 14
         |||||  |||::||||||:::||||||||||||
    1647 TACCGGAAAGCATTGGCTGTCCGAAAGATTCTACTATACTAT 1688

seq_name: gb_pl1:SCLACHXVI

seq_documentation_block:
LOCUS       SCLACHXVI     55786 bp   DNA           PLN       03-DEC-1996
DEFINITION  S.cerevisiae chromosome XVI, left arm DNA.
ACCESSION   X96770
VERSION     X96770.1  GI:1403537
KEYWORDS    BEM4 gene; CDC60 gene; KES1 gene; KIP2 gene; OYE3 gene; PA11 gene;
            PEP4 gene; PXA1 gene; REV3 gene; ribosomal protein L37a; RPL37A
            gene; SNR17B gene; spk1 gene; SVS1 gene; u3 small nuclear rna.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 55786)
AUTHORS     Purnelle,B., Coster,F. and Goffeau,A.
TITLE       The sequence of 55 kb on the left arm of yeast chromosome XVI
            identifies a small nuclear RNA, a new putative protein kinase and
            two new putative regulators
JOURNAL     Yeast 12 (14), 1483-1492 (1996)
MEDLINE     97103777
REFERENCE   2  (bases 1 to 55786)
AUTHORS     Purnelle,B.
TITLE       Direct Submission
JOURNAL     Submitted (22-MAR-1996) B. Purnelle, Unite de Biochimie
            Physiologique, Universite Catholique de Louvain, Place Croix du Sud
            2/20, 1348 Louvain-la-Neuve, BELGIUM
COMMENT     Overlapping sequences: L29279, M29683, D50278, L27816, X62878,
            Z11963, M13358, M55623, L38491, U17065, U03913, X57969, X05498.
FEATURES    Location/Qualifiers
     source  1..55786
             /organism="Saccharomyces cerevisiae"
             /strain="AB972"
             /db_xref="taxon:4932"
             /chromosome="XVI"
             /map="left arm"
             /clone_lib="pHC79 or Supercos derivative cosmids (L. Riles
             St Louis, Missouri)"
             /clone="cos9574"
             /clone="cos8484"
     gene    complement(1..542)
             /gene="OYE3"
```

```
CDS      /db_xref="SGD:S0006092"
         complement(<1. .542)
         /gene="OYE3"
         /note="L29279"
         /codon_start=1
         /db_xref="SGD:S0006092"
         /product="P2291 protein"
         /protein_id="CAA65573.1"
         /db_xref="GI:1403538"
         /db_xref="SWISS-PROT:P41816"
         /translation="MPFVKGFGEFPISLRDTNLFEPIKIGNTQLAHRAVMPPLTRMRATH
         PGNIPNKEWAAVYGQRAQRPGTMIITEGTFISPQAGGYDNAPGIWSDEQVAEWKNIF
         LAIHDCQSFAWVQLWSLGWASFPDVLARDGLRYDCASDRVYMNATIQEKAKDANNLEH
         SLTKDDIKQYIKDYIHAAKNS"

CDS      1485. .1943
         /note="putative"
         /codon_start=1
         /product="P2515 protein"
         /protein_id="CAA65551.1"
         /db_xref="GI:1403539"
         /db_xref="SPTREMBL:Q12091"
         /translation="MSFIKNLLFGGVKTSEDPTGLTGNGASNTDNSNKGSEPVVAGNF
         FPRTLSKFNGHDDEKIFIAIRGKVYDCTRGROFYGPSGPYTNFAGHDASRGLALNSFD
         LDVIKDWDQPIDPLDDLTKEQIDALDEWQEHFENKYPCIGTLIPEPGVNV"

CDS      complement(2210. .4009)
         /note="putative"
         /codon_start=1
         /product="P2520 protein"
         /protein_id="CAA65552.1"
         /db_xref="GI:1403540"
         /db_xref="SPTREMBL:Q99257"
         /translation="MSGFHNVGNINMMAQQOMQQNRIKISVRNWQNATMNDLINFISR
         NARVAVYDAHVEGPLVIGYVNSKAEAESLMKWNGVRFAGSNLKFELLDNNGASAGTSD
         TISFLRGVLLKRYDPOTKLLNLGALHSDPELIQKGVFSSISTQSKMFPAMMKLASTEK
         SLIVESVNLADNQLKDISAISTLAQTFPNLKNLCLANNQIFRFRSLEVWKNKFKDLRE
         LLMTNNPITTDKLYRTEMLRLFPKLVVLDNVIVRDEQKLQTVYSLPMKIQQFFFENDA
         LGQSSTDFATNFLNLWDNNREQLLNLYSPQSQFSVSVDSTIPPSTVTDSDQTPAFGYY
         MSSSRNISKVSSEKSIQQRLSIGQESINSIFKTLPKTKHHLQEQPNEYSMETISYPQI
         NGFVITLHGFFEETGKPELESNKKTGKNNYQKNRRYNHGYNSTSNNKLSKKSPDRTWV
         IVPMNNSVIIASDLLIVRAYSTGAWKTASIALAQPPQQQASVLPQVASMNPNITTPPQ
         PQPSVVPGGMSIPGAPQGAMVMAPTLQLPPDVQSRLNPVQLELLNKLHLETKLNAEYT
         FMLAEQSNWNIVEVAIKGFQSSMNGIPREAFVQF"

CDS      4388. .5680
         /note="putative"
         /codon_start=1
         /product="P2525 protein"
         /protein_id="CAA65553.1"
         /db_xref="GI:1403541"
         /db_xref="SPTREMBL:Q12467"
         /translation="MTVLYTSASLKKMKCLAFNMGMNCVRTVSHARSGGAKFGGRNVF
         NIFDSKTPDSVRIKAFKNTIYQSAMGKGKTKFSAMEINLITSLVRGYKGEGKNAINP
         LQTNVQILNKLLTHRLTDKDILEGMNLAAGPVNVAIPRDITPQEEKKKVELRNRKAE
         NMDLHPSRKMHIKELLHSLNLDMCNDEEVYQKISLYLQKNEESRTSVGASQQNHVDID
         INSLKRYLQNIEKKARQKSAIDKQKNQARIYQWNTQSFSEIVPLSAGNLFKREPNR
         LWKRLQNGISVFLGSNGGGKKSSKTIKKVLQGNNILLHSLENNKDMTLSNNFDHSVFNI
         NFTDLFGVINASGSPPDRVLNEINEIELKGWKCVGNLYDNNKIVVFQSSNPLLEDTKI
         PQKSPTNSKRPLISLSALLASFFAYYRYRLSQRQESKK"

gene     complement(5764. .10278)
         /gene="REV3 (also called PSO1)"
         /db_xref="SGD:S0006088"

CDS      complement(5764. .10278)
         /gene="REV3 (also called PSO1)"
         /note="M29683"
         /codon_start=1
         /db_xref="SGD:S0006088"
         /product="P2535 protein"
         /protein_id="CAA65554.1"
         /db_xref="GI:1403542"
         /db_xref="SWISS-PROT:P14284"
         /translation="MSRESNDTIQSDTVRSSSKSDYFRIQLNNQDYYMSKRPTFLDPSH
         GESLPLNQFSQVPNIVFGALPTGHVLCHVGIPLPYMFIKYDGQTDTSTLRHQRCA
         QVHKTLEVKIRASFKRKKDDKHDLAGDKLGNLNFVADVSVVKGIPFYGYHVGWNLFYK
         ISLLNPSCLSRISELIRDGKIFGKFEIYESHIPYLLQWTADFNLFGCSWINVDRCYF
```

```
         RSPVLNSILDIKLTINDLQLLLDRFCDFKCNVLSRRDFPRVGNGLIEIDILPQFIK
         NRERLQHRDIHHDFLEKLGDISDIPVKPYVSSARDMINELTMQREELSLKEYKEPPET
         KRHVSGHQWQSSGEFEAFYKKAQHKTSFDGQIPNFENFIDKNQKFSAINTPEALPQ
         LWPRLPQIEINNSMQDKKNDQVNASPTEYEICGVDNENEGVKGSNIKSRSYSWLPE
         SIASPKDSTILLDHQTKYHNTINFSMDCAMTQNMASKRKLRSSVSANKTSLLSKRKK
         VMAAGLRYGKRAFVYGEPPFGYQDILNKLEDEGFPKIDYKDPFFSNPVDLENKPYAYA
         GKRFEISSTHVSTRIPVQFGGETVSVINKPTFDMFSSWKYALAPPTYDAVQKWNKVP
         SMGNKKTESQISMHTPHSKFLYKFASDVSGKQWRKKSSVHDSLTHLTEIHANTRSDK
         IPPDAIDEVSMIIWCLEEETFPLDLDIAYEGIMIVHKASEDSTFPTKIQHCINEIPVM
         FYESEFEMFEEALIDLVLLLDPDILSGFEIHNFSWGYIIERCQKIHQFDIVRELARVKC
         QIKTKLSDTWGYAHSSGMIITGRHMINIWRALRSDVNLTQYTIESAAFNILHKRLPHF
         SPESLTNMWNAKKSTTELKTVLNYWLSRAQINIQLLRKQDYIARNIEQARLGIDFHS
         VYYRGSQFKVESFLIRICKSESFILLSPGKKDYVRKQKALECVPLVMEPESAFYKSPLI
         VLDFQSLYPSIMIGYNYCYSTMIGRVREINLTENNLGVSKFSLPRNILALLKNDVTIA
         PNGVVYAKTSVRKSTLSKMLTDLDVRVMIKTMNEIGDDNTLKRLLNNKQLALKLL
         ANYTYGTTSASFSGRMPCSDLADSIVQTGRETLEKAIDIIEKDETWNAKVVYGDTDSL
         FVLPGKTAIEAFSIGHAMAERVTQNNPKFKEKYVHPSILISKKRYVGFSYESP
         SQTLPIFDAKGIETVRRDGIPAQQKIIEKCIRLLFQTKDLSIKKYLQNEFFIQIGK
         VSAQDFCFAKEVKLGAIKSEKTAPAGAVVKKRINEDHRAEPQYKERIPYLVVKGKQG
         QLLRERCVSPEEFLEGENLELDSEYYINKILIPPLDRLFNLIGINVGNWAQEIVKSKR
         ASTTTTKVENITRVGTSATCCNCGEELTKICSLQLCDDCLEKRSTTLSFLIKKLKRQ
         KEYQILKTVCRTCSYRYTSDAGIENDHIASKCNSYDCPVFYSRVKAERYLRDNQSVQR
         EEALISLNDW"

10509. .11150
         /note="putative"
         /codon_start=1
         /product="P2540 protein"
         /protein_id="CAA65555.1"
         /db_xref="GI:1403343"
         /db_xref="SPTREMBL:Q12092"
         /translation="MINNSTNTVYIKVKGRRPQGFLDPKFEWNGTKERQLWTMVSN
         LNYSQDQIDWQNSKIFETPEFFLKKRTYKLFAEHLELLQLEKKRDLEKYSNDQVN
         EGMSDLIHKYTPLQNDNLLNVSASPLTTERQDSEEVTEVNEALQHTSKILNIH
         KKTSDSENKPNDKLDKGINKEMECGSSDDDLSSSLSVSKSALEEALMDRLQF"

complement(11127. .12248)
         /note="putative"
         /codon_start=1
         /product="P2545 protein"
         /protein_id="CAA65556.1"
         /db_xref="GI:1403544"
         /db_xref="SPTREMBL:Q12529"
         /translation="MTIDGDVHEISPFFQVRQTKWGGRACFSNGNIPKGTTVLQVSNF
         TGTSISVEFRKEVCHNCFAYANAKTMKYKLNDYLRDLVCNAHYQINPKKFLGAGLWF
         CSHHCRTSYLQIPNIIELIECYILHHFPSMLKRYNTSEQEKLNSILISENVIQS
         SWDEIESKWIPRINNMSAKRINQLPPTCEDEYCCIRFVCESLFNLKYMDPQCITYRA
         FNMLQSNELSKISKFPVLLHFQKLVFQTLYILLPSHLRMLSIPLLRHILGEYGNAF
         GLMQEGEASDSREYFGYWFPEASYFNHSCNPNITKYRKGNSMLFTMNRDIKKDEQIC
         IDYSGVLDLPTVKRRAFLADSWFFDCACERCKSELQSVH"

complement(12521. .14668)
         /note="putative"
         /codon_start=1
         /product="P2550 protein"
         /protein_id="CAA65557.1"
         /db_xref="GI:1403545"
         /db_xref="SPTREMBL:Q12083"
         /translation="MSQHIRKLDSDVSERLKSQACTVSLASAVREIVQNSVDAHATTI
         DVMIDLPNLSFAVYDDGIGLTRSDLNILATQNYTSKIRKMNDLVTMKTYGYRGDALYS
         ISNVSNLFVCSKKKDYNSAWMRKFPSKSVMLSENTILPIDPFWKICPWSRTKSGTVVI
         VEDMLYNLPVRRILKEEPPFKFNTIKADMLQILVMHPMISLNVQYTDKLRINTEVL
         FRSKNITEGLRKHQQMSQVLRNVFGAIIPPDMLKKVSLKFNEYQIEGIISKMPVGLKD
         LQFIYINGRRYADSAFQGYVDSLFQAQDFGEKGMSLLKTKSVGKYRSHPVFILDVRC
         PQTIDDLLQDPAKSIAVFQSYRHIRTIEPLIVKTIRSFLITFQGYLTPDKSDSSFEIVNCSQ
         KTATLPDSRIQISKRNQVLNSKMKIARINSYIGKPAVNGCRINNSTINYEKIKNIRID
         GQKSRLRNKLSSRPYDSGFTEDYDSIGKTIIDFSISRSVLAKYEVINQVDKKFILIRC
         LDQSIHNCPLLVVDVQHACDERIRLEELFYSLLTEVVTGTFVARDLKDCCIEVDRTEA
         DLFKHYQSEFKKWGIGYETIEGTMETSLLEIKTLPEMLTSKYNGDKDYLKMVLLQHAH
         QECIILISLKSRCHNPFFCAHGRPSMVPIAELR"

gene     complement(15089. .15871)
         /gene="SVS1"
         /db_xref="SGD:S0006084"

CDS      complement(15089. .15871)
         /gene="SVS1"
```

```
                1 TyrArgLeuAlaArgIle***ArgIleLeuArgTyr 14
                  ||||||::::||    |||::::  :::  ||||||||||||
             3148 TATAGGATAGCTGAAAGATTCGAAAAATGCTACTTCGAPAT 3189

seq_name: gb_inl:CEE02A10

seq_documentation_block:
LOCUS       CEE02A10   23889 bp    DNA              INV      02-SEP-1999
DEFINITION  Caenorhabditis elegans cosmid E02A10, complete sequence.
ACCESSION   Z81053
VERSION     Z81053.1  GI:1729531
KEYWORDS    HTG: Calmodulin-like protein; Mitochondrial ribosomal protein S5.
SOURCE      Caenorhabditis elegans
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 23889)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
REFERENCE   2  (bases 1 to 23889)
AUTHORS     Thomas,K.
TITLE       Direct Submission
JOURNAL     Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Dec 12, 1996 this sequence version replaced gi:1627715.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormaceclass=Sequence sobject=E02A10
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone E02A10.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C14C10 is at 23796 in this sequence. The
            true right end of clone F45D3 is at 107 in this sequence. The start
            of this sequence (1..108) overlaps with the end of sequence 278063.
            The end of this sequence (23793..23889) overlaps with the start of
            sequence 274028.
FEATURES             Location/Qualifiers
     source          1..23889
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="v"
                     /clone="E02A10"
     gene            complement(11007..12594)
                     /gene="E02A10.2"
     CDS             complement(join(11007..11504,11551..11813,11984..12016,
```

```
                     /note="D50278"
                     /codon_start=1
                     /db_xref="SGD:S0006084"
                     /product="P2554 protein"
                     /protein_id="CAA65558.1"
                     /db_xref="GI:1403546"
                     /db_xref="SWISS-PROT:Q12254"
                     /translation="MIFKILCSLLLVTSNFASALYVNETTSVTPYTKTLTPTYSVSPQ
                     ETTLFYSDETTFYIITSTFYSTYWFTTSQSAAIISTPTASTPTASTPSLTTSTNEYTI
                     TYSDTDTYTSLITSTYIITLSTESANEKAEQISTSVTEIASTVTESGSTYTSTLTST
                     LIVTVYNSQASNIATSTAGDRAASNVDALEKLVSAEHQSQMIQTTSADEQYCSASTKY
                     VTVTAAAVTEVVTTAEPVKYVTITADASNVTGSANNGTHI"
CDS              complement(16377..17198)
                     /note="putative"
                     /codon_start=1

alignment_scores:
       Quality:     41.00         Length:      14
       Ratio:        3.727        Gaps:         0
Percent Similarity:  78.571     Percent Identity:  57.143

alignment_block:
US-08-653-294-20 x SCLACHXVI/rev    .

Align seq 1/1  to reverse of: SCLACHXVI  from: 1  to: 55786

                1 TyrArgLeuAlaArgIle***ArgIleLeuArgTyr 14
                  ||||||  |||::::|||:  ::||||||||| |||
             8936 TACCCGAAAGCATTGCGTCTCGAAAGATTCTACTATACTAT 8895

seq_name: gb_htg5:AC013211

seq_documentation_block:
LOCUS       AC013211    3727 bp    DNA         HTG        03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013211
VERSION     AC013211.1  GI:6223121
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3727)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214157 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..3727
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT      977 a    847 c    880 g   1023 t
ORIGIN

alignment_scores:
       Quality:     40.00         Length:      14
       Ratio:        3.636        Gaps:         0
Percent Similarity:  78.571     Percent Identity:  57.143

alignment_block:
US-08-653-294-20 x AC013211    .

Align seq 1/1  to: AC013211  from: 1  to: 3727
```

```
seq_documentation_block:
LOCUS       CEM01F1          36355 bp    DNA                INV       02-SEP-1999
DEFINITION  Caenorhabditis elegans cosmid M01F1, complete sequence.
ACCESSION   Z46381
VERSION     Z46381.1  GI:561920
KEYWORDS    HTG; Glucose transport protein; L13P ribosomal protein; Lipoic acid
            synthase; Ysy6.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 36355)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 36355)
  AUTHORS   Sims,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1994) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence&object=M01F1
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone M01F1.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone M01F1 is at 1 in this sequence. The
            start of this sequence (1. .104) overlaps with the end of sequence
            Z46996.
            The end of this sequence (36252. .36355) overlaps with the start of
            sequence Z77131.
FEATURES             Location/Qualifiers
     source          1..36355
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="M01F1"
     gene            complement(3577. .6382)
                     /gene="M01F1.1"
     CDS             complement(join(3577. .3690,3776. .3845,3899. .4008,
                     4240. .4330,4518. .4667,4911. .4990,5046. .5155,5277. .5533,
                     5651. .5706,5757. .5807,6111. .6251,6299. .6382))
                     /gene="M01F1.1"
                     /note="similar to alpha-1,3-mannosyl-glycoprotein
                     beta-1,2-N-acetylglucosaminyltransferase"
                     /codon_start=1
                     /protein_id="CAA86513.1"
```

```
            12064. .12174,12220. .12258,12324. .12443,12501. .12594))
            /gene="E02A10.2"
            /note="predicted using Genefinder"
            /codon_start=1
            /protein_id="CAB02877.1"
            /db_xref="GI:387541"
            /db_xref="SPTREMBL:Q93424"
            /translation="MRKSIIILLLGLSVATAFLFPSGGGGGGGGGGGGGGGGGGGGCGG
            GGGCGGGCAPPPPPPACGGCGCGGGGGGGGGCGGGGCGGGGGGGGGGGGGGGGCGGGGGG
            GGCGGGGGGGGGGGCCGGGCGGGGCGGGGGGGRKKREAINVGHDDLNKCNNEELRIIL
            NENTKETVAESIKTLKEKVAGQDYIVVCNEKPAPFTAETDDFCSLQKENVHCTLRIN
            HKEVAEKNEDKKEEPKKEEKKEVEKKEEDEKKDEPKKEEKKEEQKEEVEKK
            EEEEKKDEPKKEEKKKEEKKEEEKKEDEVEEKSEKVEKELEPKKDEETKKN"
gene        22553. .23644
            /gene="E02A10.3"
CDS         join(22553. .22577,22896. .22982,23039. .23136,23285. .23644)
            /gene="E02A10.3"
            /note="predicted using Genefinder; Similarity to C.elegans
            calmodulins"
            /codon_start=1
            /protein_id="CAB02878.1"
            /db_xref="GI:387542"
            /db_xref="SPTREMBL:P90802"
            /translation="MTNSVFYSDPLLSQHLLEGYSEEELQEYRQVFNMFDADRSGAIA
            IDELEAAIKNLGLEQTRDELDKIIDEVDQRGNHQJDFDEFCVVMRRLTMKKSNWNEVF
            KECFTVFDRVSSNACRFCDSNSFVFQSENGGISKKDFRFILRELGDITDNQIIDEIFN
            EADVDGNGVIDYDEFTYMYKNYMIDDDIY"
gene        join(278063.1:36566. .36692,278063.1:36740. .36890,
            278063.1:36943. .37434,109. .194,841. .947,998. .1132,
            1180. .1338)
            /gene="E02A10.1"
CDS         join(278063.1:36566. .36692,278063.1:36740. .36890,
            278063.1:36943. .37434,109. .194,841. .947,998. .1132,
            1180. .1338)
            /gene="E02A10.1"
            /note="predicted using Genefinder; Similarity to Yeast
            mitochondrial ribosomal protein S5 (SW:RT05_YEAST); cDNA
            EST EMBL:D65461 comes from this gene; cDNA EST EMBL:D68901
            comes from this gene; cDNA EST yk432a4.3 comes from this
            gene; cDNA EST yk432a4.5 comes from this gene; cDNA EST
            yk554d6.3 comes from this gene"
            /codon_start=1
            /protein_id="CAB02879.1"
            /db_xref="GI:387543"
            /translation="MRRSGPELWKTLTSVSKSGQKKGRRNTRQPVRPLNRFYRIGSSP
            MKIFAGLNAPIRMRETENQNLMSAEQTEDEIRDSMGGTKKILEERDTGKKKRNREK
            LHPMERGFSGTQLVGQKLGAPPPLDGVNFDDFETYCLEVKRTSNMTNVFGRVHTMSAL
            VVTGNGRGLAGYAVGKAPIHRTTAIINGMGMASRKLFHVELHEGRTIYQDFYAECRN
            TRVFAQRRPRGFGLTCPHRLIKCEAIGIKDIYVVEGSKNYLALTHAFVTGLLNQE
            THQQLAERKGLHVVEMSPSRHFLPQIVASPISTELKTEETLEALDRLNLDDFYGEGRY
            PLRKPKSLPFFSNLEGHLDARWRKHPFRNQESTMIRLIADNMVPRWTRDARAAWADQR
            NERMTTGVPMPLGLGLSHVVPKKDD"
BASE COUNT   7821 a   4089 c   3910 g   8069 t
ORIGIN
```

alignment_scores:
```
Quality:        40.00         Length:        13
Ratio:          3.636         Gaps:           0
Percent Similarity: 84.615    Percent Identity:  61.538
```

alignment_block:
```
US-08-653-294-20 x CEE02A10
Align seg 1/1  to: CEE02A10  from:  1  to: 23889

     1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArg 13
       :::::||||||::||||| ||:::||||:|:|||||||
 11706 CACCGGCCTCCTCCCTCCCCATCCTCCGCCTCCTCCTCCGC 11744
```

seq_name: gb_in|:CEM01F1

/db_xref="GI:3878567"
/db_xref="SPTREMBL:Q21450"
/translation="MHISSKITFIFVLIYFSWNFYIQNGFTNKAIVKEELRLLKEMDE
AGSRLEKLLKVEKEHMERLSEIMNRATRLRITPNSNNPNNWPHPIVIVFSCNRPDSV
RAHVEKLIMYRPSAQQFPITVSODCDNESVKKEVEKFGSVNYVKHPPGESVKIEIPT
NLQKFKPYYISRHYKLALNHIFSNSNNYSSVIITEDDLDIAPDFFSYFSNTRYLLEK
DPSLWCVTAWNDNGKPENIDLKSNATLYRSDFFAGLGWMMIRKTWELEPIWPNGFWD
DWMREPVQRKQRCIRPEISRTGMMKYGKEGTSKGQFSDHLEKIKNDLPVDFSQIN
LDYLQKNEFESRLSLDIRNAVPVDIDDITYPDWKPDYEGMKAIIYYTGRTDFVAKADR
LSLMHDFKAGVPRTAYNGIVTCFYKGTRIFLVPDRSKVPGYDSSW"
/gene="M01F1.8"
complement(12635..15694)
/gene="M01F1.8"

gene

CDS
complement(join(12635..12768,13939..14239,15031..15225,
15281..15548,15553..15694))
/gene="M01F1.8"
/note="cDNA EST yk415a9.3 comes from this gene; cDNA EST
yk415a9.5 comes from this gene; cDNA EST yk463b4.3 comes
from this gene"
/codon_start=1
/protein_id="CAA86514.1"
/db_xref="GI:3878568"
/db_xref="SPTREMBL:Q21451"
/translation="MATILHKGGSVVAVSHDMMVRYDGKGVLFKKELLTRGYKHDDI
VQFNARKATTVDIEVGDKLSOMANYELAWVGTELKFIKQGRMLVAEGAVINIPDDKAQ
YMFVDWKYTKNPPTLAKVCGNIPGYIKPLTINVFVTRATVRPDVEKLTEMFKPGQRVK
FVAREQAPNERGCWRAALATDEYHDIAIDVPNAGRQTYRVHPKAGVCPPAPLQQRP
LAPPNPQVPSTSAWNIAPTPTPVTVQRPLRKPVTPQPVKSAFQAQKPPPIDGFDYFDS
SKEACAARFRNFANTIRKNNPNCIFSQPCFEGALKIKYLNQIHEELTVMREGEVDY"
complement(16047..16754)
/gene="M01F1.2"

gene

CDS
complement(join(16047..16243,16291..16443,16496..16754))
/gene="M01F1.2"
/note="similar to L13P family ribosomal protein; cDNA EST
EMBL:M79788 comes from this gene; cDNA EST EMBL:D71976
comes from this gene; cDNA EST EMBL:D72020 comes from this
gene; cDNA EST EMBL:D72496 comes from this gene; cDNA EST
EMBL:D66294 comes from this gene; cDNA EST EMBL:D73026
comes from this gene; cDNA EST EMBL:D75365 comes from this
gene; cDNA EST EMBL:D75395 comes from this gene; cDNA EST
EMBL:D74801 comes from this gene; cDNA EST EMBL:D74740
comes from this gene; cDNA EST EMBL:D70026 comes from this
gene; cDNA EST EMBL:C08236 comes from this gene; cDNA EST
EMBL:C09924 comes from this gene; cDNA EST EMBL:C08172
comes from this gene; cDNA EST EMBL:C07790 comes from this
gene; cDNA EST EMBL:C09296 comes from this gene; cDNA EST
EMBL:C07403 comes from this gene; cDNA EST EMBL:C08764
comes from this gene; cDNA EST EMBL:C07707 comes from this
gene; cDNA EST EMBL:C09161 comes from this gene; cDNA EST
EMBL:C07653 comes from this gene; cDNA EST EMBL:C09085
comes from this gene; cDNA EST EMBL:C07577 comes from this
gene; cDNA EST EMBL:C08961 comes from this gene; cDNA EST
EMBL:C10177 comes from this gene; cDNA EST EMBL:C11717
comes from this gene; cDNA EST yk503e5.3 comes from this
gene; cDNA EST yk500g9.3 comes from this gene; cDNA EST
yk486g2.3 comes from this gene; cDNA EST yk486g2.5 comes
from this gene; cDNA EST yk475e7.3 comes from this gene;
cDNA EST yk475e7.5 comes from this gene; cDNA EST
yk466e4.3 comes from this gene; cDNA EST yk466e4.5 comes
from this gene; cDNA EST yk44868.3 comes from this gene;
cDNA EST yk44868.5 comes from this gene; cDNA EST
yk432f11.3 comes from this gene; cDNA EST yk432f11.5 comes
from this gene; cDNA EST yk42e74.3 comes from this gene;
cDNA EST yk42e74.5 comes from this gene; cDNA EST
yk394a8.3 comes from this gene; cDNA EST yk394a8.5 comes
from this gene; cDNA EST yk400a3.3 comes from this gene;
cDNA EST yk400a3.5 comes from this gene; cDNA EST
yk374d9.3 comes from this gene; cDNA EST yk374d9.5 comes
from this gene; cDNA EST yk391h8.3 comes from this gene;
cDNA EST yk391h8.5 comes from this gene; cDNA EST
yk372g8.3 comes from this gene; cDNA EST yk372g8.5 comes
from this gene; cDNA EST yk372b4.3 comes from this gene;
cDNA EST yk372b4.5 comes from this gene; cDNA EST
yk362c1.3 comes from this gene; cDNA EST yk362c1.5 comes

from this gene; cDNA EST yk298a4.3 comes from this
gene; cDNA EST yk298a4.5 comes from this gene; cDNA EST
yk330b10.3 comes from this gene; cDNA EST yk330b10.5 comes
from this gene; cDNA EST yk327c5.3 comes from this gene;
cDNA EST yk327c5.5 comes from this gene; cDNA EST
yk300b7.3 comes from this gene; cDNA EST yk300b7.5 comes
from this gene; cDNA EST yk313a3.3 comes from this gene;
cDNA EST yk313a3.5 comes from this gene; cDNA EST
yk252e3.3 comes from this gene; cDNA EST yk252e3.5 comes
from this gene; cDNA EST yk235h1.3 comes from this gene;
cDNA EST yk235h1.5 comes from this gene; cDNA EST
yk235b11.3 comes from this gene; cDNA EST yk235b11.5 comes
from this gene; cDNA EST yk234f10.3 comes from this gene;
cDNA EST yk234f10.5 comes from this gene; cDNA EST
yk211a2.3 comes from this gene; cDNA EST yk211a2.5 comes
from this gene; cDNA EST yk498h1.3 comes from this gene;
cDNA EST yk508f12.3 comes from this gene; cDNA EST
yk518c5.3 comes from this gene; cDNA EST yk538e5.3 comes
from this gene; cDNA EST yk570f5.3 comes from this gene;
cDNA EST yk573b7.3 comes from this gene; cDNA EST
yk587f2.3 comes from this gene; cDNA EST yk589f3.3 comes
from this gene; cDNA EST yk599c7.3 comes from this gene;
cDNA EST yk606h9.3 comes from this gene; cDNA EST
yk609e8.3 comes from this gene; cDNA EST yk623g5.3 comes
from this gene; cDNA EST yk633f11.3 comes from this gene;
cDNA EST yk649f12.3 comes from this gene; cDNA EST
yk660e11.3 comes from this gene; cDNA EST yk670f1.3 comes
from this gene"
/codon_start=1
/protein_id="CAA86515.1"
/db_xref="GI:3878569"
/db_xref="SWISS-PROT:Q27389"
/translation="MGLSNRAIIIDGKNHLLGRLASIVAKKLLQGDKVVLRAEEIYI
SGNFHRSKLKIMSFLRKRCNINPARGAFHYRAPGKIFWRTVRGMLPHKTNRGNEALKN
LRAYEGVPAKYQKTKSLHAPSASRFRLQPRRKFCVVGRLSHEVGWQFQDVVAKLEAKR
KVKGAAYFEQKKKMDKLAVQAKKNAAPKIAQYQKIIEALGYN"
complement(16937..18194)
/gene="M01F1.3"
complement(join(16937..17265,17322..17557,17602..17903,
17948..18045,18095..18194))
/gene="M01F1.3"
/note="similar to lipoic acid synthase; cDNA EST yk283b6.3
comes from this gene; cDNA EST yk283b6.5 comes from this
gene; cDNA EST yk47f5.3 comes from this gene; cDNA EST
yk47f5.5 comes from this gene; cDNA EST yk476e7.3 comes
from this gene; cDNA EST yk476e7.5 comes from this gene;
cDNA EST yk543a1.3 comes from this gene; cDNA EST
yk588g11.3 comes from this gene; cDNA EST yk662d9.3 comes
from this gene"
/codon_start=1
/protein_id="CAA86516.1"
/db_xref="GI:3878570"
/db_xref="SPTREMBL:Q21452"
/translation="MLARVWVRGLAATKKKPQVLVKDGPSLQDFISSASVAEAVEKYE
GKLKLEKGDRRLRLPWLKKEKILPSENENVSRLKKQLKHLKLATVCQEARCPNLGEC

gene

CDS

alignment_scores:
　　Quality:　　40.00　　　　Length:　　13
　　　Ratio:　　3.636　　　　　Gaps:　　0
Percent Similarity:　84.615　　Percent Identity:　69.231

alignment_block:
US-08-653-294-20 x CEM01F1/rev

Align seg 1/1　to reverse of: CEM01F1　from: 1　to: 36355

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArg 13
      ::::::|||　|||||||||　|||||||||||||
 4482 TTTGAGCTGAAATTCGAATTCTGCGCATTTTGTTACGT 4444

seq_name= gb_inl:CET21C9

seq_documentation_block:

```
LOCUS        CET21C9      37545 bp   DNA              INV        02-SEP-1999
DEFINITION   Caenorhabditis elegans cosmid T21C9, complete sequence.
ACCESSION    Z73098
VERSION      Z73098.1  GI:1313944
KEYWORDS     HTG; Discs large; GTP cyclohydrolase I; KIP1; Myosin heavy chain
             IB; pyruvate, water dikinase: Sodium/chlorine dependent GABA
             transporter.
SOURCE       Caenorhabditis elegans.
  ORGANISM   Caenorhabditis elegans
             Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
             Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE    1  (bases 1 to 37545)
  AUTHORS    Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
             Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
             Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
             Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier.M.,
             Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
             Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
             O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
             Saunders,D., Showknen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
             Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
             Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
             Wilkinson-Sproat,J. and Wohldman,P.
  TITLE      2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
             elegans
  JOURNAL    Nature 368 (6466), 32-38 (1994)
  MEDLINE    94150718
REFERENCE    2  (bases 1 to 37545)
  AUTHORS    McMurray,A.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-MAY-1996) Louis, MO 63110, USA. E-mail:
             jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT      Coding sequences below are predicted from computer analysis, using
             predictions from Genefinder (P. Green, U. Washington), and other
             available information.
             For a graphical representation of this sequence and its analysis
             see:-
             http://webace.sanger.ac.uk/cgi-
             bin/display?db=wormace&class=Sequence &object=T21C9
             Current sequence finishing criteria for the C. elegans genome
             sequencing consortium are that all bases are either sequenced
             unambiguously on both strands, or on a single strand with both a
             dye primer and dye terminator reaction, from distinct subclones.
             Exceptions are indicated by an explicit note.
             IMPORTANT: This sequence is NOT necessarily the entire insert of
             the specified clone. It may be shorter because we only sequence
             overlapping sections once, or longer because we arrange for a small
             overlap between neighbouring submissions.
             This sequence is the entire insert of clone T21C9. The true right
             end of clone F32G8 is at 17256 in this sequence. The start of this
             sequence (1 ..105) overlaps with the end of sequence Z72509.
             The end of this sequence (37442 ..37545) overlaps with the start of
             sequence Z95392.
FEATURES             Location/Qualifiers
     source          1..37545
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="T21C9"
     gene            complement(446..10197)
                     /gene="T21C9.2"
     CDS             complement(join(446..548,595..1000,1024..1238,1292..2886,
                     2934..3086,4810..4914,4965..5145,9236..9475,9998..10197))
                     /gene="T21C9.2"
                     /note="Weak similarity to the yeast KIP1 protein (Swiss
                     Prot accession number P28742); cDNA EST CEMSF39F comes
                     from this gene; cDNA EST EMBL:D34039 comes from this gene;
                     cDNA EST EMBL:D34440 comes from this gene; cDNA EST
                     EMBL:D37039 comes from this gene; cDNA EST EMBL:D37503
                     comes from this gene; cDNA EST EMBL:D64420 comes from this
                     gene; cDNA EST EMBL:D64863 comes from this gene; cDNA EST
                     EMBL:D67399 comes from this gene; cDNA EST EMBL:D68114
                     comes from this gene; cDNA EST EMBL:D68330 comes from this
```

```
                     gene; cDNA EST EMBL:C10882 comes from this gene; cDNA EST
                     EMBL:C12782 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA97331.1"
                     /db_xref="GI:3880015"
                     /db_xref="SPTREMBL:Q22639"
                     /translation="MKSMDDRTVGLISASSSRASLKIQPEIGYPRICDYCQPIIELMT
                     ASEFARHIRQDHTTKEGGSFLCRYGEHGVCQKLPLEGVCDLDFAHIRCHTSSOPAG
                     YSSPTSSLVASYTEDSEETASLRSIRLTSDRDTPTIEKKKFTLHSFTQNLSAVLADPS
                     RSRSDLTTFFTRHWGDSFVPTQPVPQSKRIARLADSAFDSYCQSAGESYRRYQAIKRA
                     LRLSHADGGVGIGDERODAEDLPTFIDPRFTLGDSTFSDVFTVPANDDLDALKQTL
                     SGRNVIPATPLEIATNRKPGEFRDYEALQNRLEMMHDVVDGRLAGKLVAKTDDFWQVV
                     RSYSGLQEQANALQCVMVRKNLKRVDELVCDQSKKIVEVHERYEQKKLLAKLHDI
                     SCLREAQSTVQMLSQGDYPKAIECIETSLDVLSKELNGVTCFRHLASQLRELYIVIG
                     RMNNEDFASLIQKELGVKPEAGTMIQAEGELSAVLLGLMRMKYSFIAVLREEIEGV
                     KSVMRQVIKNQILNNGVDLNDFDPSLTQLGEPVRRMKHADFLKIVKAVMDEEFYFCKR
                     LEALQDILLETVERANPLHRHGSEDIIVERIEEAKDIHESESDDEVGTTSKSASSGG
                     FSVGGSALSSNASATTLLSIEVRSEAFLRHVLPLIAEFGHQCAQQRISRLLIARAKNA
                     SVTEATTPTQLSECIKLVKEFQSKCDKEGWYSTQNQKVGGLGRSVNKLSMDYIEKFHA
                     ARKIRIGNMLDTELWKATDVSIVDONIVDMAMETGQLRNEI AAKLISVIAHYTNCLGMDVKGVIP
                     VTIDSTTSTSNQIQGVIVDEENYVVVGSSTIMIQLLSDYCEAISEMPSFSQDWNSRVV
                     ELLKTFNSRCCQLIIGAGALQLVGLKTISVRNLGIALKFYPLAGRSLELVCRFIPMVH
                     DEMDRVLPENRKSLLRYFKQVESEYRDHVNEIAAKLISVIAHYTTNCLGWDVKGVIP
                     SPEFQQICRHMLKFHNGLVGIMPRDQIEALFRQVHENFKANLREHVTGMGITPHDPLK
                     YGYVTKDYMFYQQNVKNMESCRNLELESLNDIMFE"
     gene            complement(10754..12112)
                     /gene="T21C9.1"
     CDS             complement(join(10754..10853,11143..11271,11317..11594,
                     11645..11775,11821..11928,12049..12112))
                     /gene="T21C9.1"
                     /note="Similarity to the drosophila discs large protein
                     repeats; cDNA EST EMBL:T01372 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA97330.2"
                     /db_xref="GI:584627"
                     /db_xref="SPTREMBL:Q22638"
                     /translation="MMVSPPQEDTVFNTDSDPVEQATDDMTFNDTTDGNQESVPL
                     EALTVEIEKTSKGFNIVGGTDNPHFVGDIGIYVSSVNSESKSYGVVRTGDKILSF
                     DGIDMTYKTHDEAVEVFRSVKIGHVAKMLIDREYLHLQEDRTQPTASVSITPQVTPQ
                     TRSTQNNTDTPKSMSHSESKSRLTSHGLSAVIERIGRGKYEEEAQSVTSYAPSTHSI
                     IDDVPRTPRKPLSLLDPRNNSWLEALYVSIGLGALTISGYLAYRFIRGRR"
     complement(12390..12817)
                     /gene="T21C9.4"
     complement(join(12390..12479,12525..12655,12727..12817))
                     /gene="T21C9.4"
                     /note="Similarity to drosophila hypothetical protein PIR
                     accession number SS0341; cDNA EST yk391d9.3 comes from
                     this gene"
                     /codon_start=1
                     /protein_id="CAA97332.1"
                     /db_xref="GI:3880016"
                     /db_xref="SWISS-PROT:Q22640"
                     /translation="MSHTILLLQPIDNIESRSWSDYENTTECLEGICRVYEEYLKKKV
                     PAQNEITYDISHLFEFIDDLKDLSMLVLDNTTTYVPHNKQYVKESIYKLMNNRLNNQ
                     H"
                     13113..14343
                     /gene="T21C9.5"
                     join(13113..13183,13598..13730,14152..14343)
                     /gene="T21C9.5"
                     /note="cDNA EST EMBL:T01297 comes from this gene; cDNA EST
                     yk257f11.5 comes from this gene; cDNA EST yk257f11.5 comes
                     from this gene"
                     /codon_start=1
                     /protein_id="CAA97333.1"
                     /db_xref="GI:3880017"
                     /db_xref="SPTREMBL:Q22641"
                     /translation="MSGLSHAGRQVARIAVRQASSHSHDSHAVWKEINRLGSEGKWDN
                     VNNMPKMFLFGEAKQETTAYRAINKDPDFFRQSPYGQYLKIVWRLALLFGIIKAGTV
                     VYFAVPEEQRLKYKYRNHGHHGHDDAHD"
                     15683..20825
                     /gene="T21C9.6"
                     join(15683..15783,16328..16412,16458..16849,16892..16923,
                     17029..17093,17175..17660,17707..17846,17897..18136,
```

     gene

     CDS

     gene

     CDS

     gene

     CDS

```
DEFINITION   Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MZN24,
             complete sequence.
ACCESSION    AB028622
VERSION      AB028622.1  GI:5041975
KEYWORDS     HTG.
SOURCE       Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
             clone:MZN24.
ORGANISM     Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
             eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
             Arabidopsis.
REFERENCE    1 (sites)
AUTHORS      Nakamura,Y.
TITLE        Structural Analysis of Arabidopsis thaliana Chromosome 3. II
JOURNAL      Unpublished (1999)
REFERENCE    2 (bases 1 to 82348)
AUTHORS      Nakamura,Y.
TITLE        Direct Submission
JOURNAL      Submitted (09-JUN-1999) to the DDBJ/EMBL/GenBank databases.
             Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
             Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
             (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
             Fax:+81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..82348
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="MZN24"
                     /clone_lib="Mitsui P1"
BASE COUNT     27062 a  13968 c  13880 g  27438 t
ORIGIN
```

alignment_scores:
    Quality: 40.00   Length: 14
    Ratio: 4.000   Gaps: 0
Percent Similarity: 71.429   Percent Identity: 50.000

alignment_block:
US-08-653-294-20 x AB028622   ..

Align seg 1/1 to: AB028622 from: 1 to: 82348

      1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
        :::|||||| ::::||| |||:::: ||||||||
  32984 TTCAGGTGATAGTGCGAACGTATCGTGTGACTCTAAGATAT 33025

seq_name= gb_p12:ATAC007070

seq_documentation_block:
LOCUS   ATAC007070   86017 bp   DNA   PLN   06-APR-1999
DEFINITION Arabidopsis thaliana chromosome II BAC T22F11 genomic sequence,
            complete sequence.
ACCESSION   AC007070
VERSION   AC007070.3 GI:4567237
KEYWORDS   HTG.
SOURCE   thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1 (bases 1 to 86017)
AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
            Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
            Fraser,C.M. and Venter,J.C.
TITLE   Arabidopsis thaliana chromosome II BAC T22F11 genomic sequence
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 86017)
```

```
18184..18403,18484..18801,18852..19332,19645..20053,
20104..20239,20283..20825)
/gene="T21C9.6"
/note="similar to pyruvate, water dikinase; cDNA EST
EMBL:D64535 comes from this gene; cDNA EST EMBL:D67561
comes from this gene; cDNA EST EMBL:D67645 comes from this
gene; cDNA EST yk350e8.3 comes from this gene; cDNA EST
yk350e8.5 comes from this gene; cDNA EST yk397f12.3 comes
from this gene; cDNA EST yk397f12.5 comes from this gene;
cDNA EST yk295f7.3 comes from this gene; cDNA EST
yk295f7.5 comes from this gene; cDNA EST yk227e6.5 comes
from this gene; cDNA EST yk271d10.5 comes from this gene"
/codon_start=1
/protein_id="CAA97341.1"
/db_xref="GI:3880025"
/db_xref="SPTREMBL:Q22649"
/translation="MLIFLFLLSAPLCKFVLYPVIYAFLKWRHPVKKFTAPTAKIR
DIINPDKEEWELEHIKNEEILTIGSGGEDNSFVYIQVITKNEIHFAHIRVFEKGILYS
GSYKAVFSDNRTISCGPLLIELRNPFRKWRINFRGYLNDAKGNSHFVILSGWWRCVSN
ARFFYSNALLKSFNEIYNDKPTALIENLKIIEKYKKGNVLNQMGLRHRNPLDLETSST
HFHTYLSDANAISYRVNSVDEKKSVSHGVGFRSDHSARPIIIDNSHSLESDNMLPVQF
KYVSRGLLPLEMRKGRKLCHFTFIREDNKVVDVTALEVNSAIFTGVGFVVRVREAIHE
VTTKLPDSFPEYHATSEKMKSVVSFNHRACQDKMLTGGKGANLARLQAITDDFHVPP
GIVVTTAAFNKHVIANPNVLEEIKLLDINDKNAEYYEDVGKRIEGLLFESEVSQELQK
EIKEWLPRSEYAVRSSAVGEDGADLSSAGQLESLDVIEIDIVDKLKLCWGSNFRRE
VLNYRKNYGQQLNPSMAVVINKVGAGVMFTANPVKLDRGEIVINALKGSGEGVV
SGVTTPDEIYVNRIHKTVVINKVGVDCCLDDYQIEKLTKVGEYLERIFGKPQDIEFVV
RNQMVNIVQSRDITGLDKETQFEMCTEYNSPSIHDKEILINANVGEVLPVVNAMEAH
NLTGMFDKVIASMTFQELNDVVPAHISIGFSVAHRIFFNLGEVVLRIWELVEKDRIT
DIVIAGETLFNAEMFQQAAHQYGKVYSPFFPLKRMFNMIKLIYFTSNSVKAKIVKIDEE
AKKLVPNDKMTIEIVFKKYDELEKLWCDASHCHTSLSMFSSFTVLCGMLIRGSDNGP
LSNDNISDFANVFSNNSRGDVISADVPNSLKKLAKTIREESIDKEFCAAEDDEAIKTL
ENGKKSSDELRRFLELHGHRGPKELYLDATTWEEDTNLVHTIKSMLACPETSGKTIE
NEDDIIDNLKCKPTGIRRRLLKYFIGOTHRGVSFRETARNHLVSTHSLRKACRMVGK
KLYEKGYLPDPNLWMHFSMDELKELNATRSAKLVSRAIRRKQISSKFEGLQFPLVSHG
YMNPIKIEIAETDANIGILLRGTTVCEGKVRARARVAKTLEEAKETKPGEILITKYTD
ICWSPFFPIISGIVTEIGGLLSHGAVVAREYGLPSLIAVSNATYHFKTGDLVELDSIS
GTISRDTNQE"
gene         complement(21064..22585)
             /gene="T21C9.7"
CDS          complement(join(21064..21207,21256..21444,21587..21724,
             21806..22039,22442..22585))
             /gene="T21C9.7"
             /codon_start=1
             /protein_id="CAA97334.1"
             /db_xref="GI:3880018"
             /db_xref="SPTREMBL:Q22642"
             /translation="MLTPLFFLPAAYGVASSYLYTLIILMMIRRWNEYNTAFFKLFII
EYVFNMVTFANSFVLRAFQNTCNNCTFAFLFERNSSTEEDNFPLQVFFIHYCMAFI
QYFMTFLSSLNRLIMVFDSYEKIWRPSLPFLIMVIIFPMIMTWPIATNNAYFIYS
PTLGGFATKTVANSTEVLNSLVTFMVFTLFTATANIVSIRLTLLPTRISGAERNLF
TVSFVSFIIQLLALGDTLILRAAPSEGMSVGAQTAKALMPFVSDILTFNHPWTIMYFS
TKVNDI"
gene         complement(23002..23649)
```

alignment_scores:
    Quality: 40.00   Length: 14
    Ratio: 4.000   Gaps: 0
Percent Similarity: 71.429   Percent Identity: 57.143

alignment_block:
US-08-653-294-20 x CET21C9   ..

Align seg 1/1 to: CET21C9 from: 1 to: 37545

      1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
        :::|||||| ::::||| |||:::: ||||||||
  11656 TATCGCCTGTTCTTACGACTCCATATGATTTTGACTCGCTAT 11697

seq_name= gb_p11:AB028622

seq_documentation_block:
LOCUS   AB028622   82348 bp   DNA   PLN   20-NOV-1999

```
AUTHORS   Lin,X. and Kaul,S.
TITLE     Direct Submission
JOURNAL   Submitted (12-MAR-1999) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 3  (bases 1 to 86017)
AUTHORS   Lin,X.
TITLE     Direct Submission
JOURNAL   Submitted (06-APR-1999) The Institute for Genomic Research, 9712
          Medical Center Dr., Rockville, MD 20850, USA
COMMENT   On Apr 6, 1999 this sequence version replaced gi:4454441.
          Address all correspondence to:
          Xiaoying Lin
          The Institute for Genomic Research
          9712 Medical Center Dr.
          Rockville, MD 20850, USA
          e-mail: xlin@tigr.org
          BAC clone T22F11 is from Arabidopsis chromosome II and is near the
          molecular marker GPA1.
          The orientation of the sequence is from SP6 to T7 end of the BAC
          clone.
          Genes were identified by a combination of three methods: Gene
          prediction programs including GRAIL (available by anonymous ftp
          from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
          Washington), Genscan (Chris Burge),
          http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
          (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
          complete sequence against a peptide database and the Arabidopsis
          EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
          Annotated genes are named to indicate the level of evidence for
          their annotation. Genes with similarity to other proteins are named
          after the database hits. Genes without significant peptide
          similarity but with EST similarity are named as 'unknown' proteins.
          Genes without protein or EST similarity, that are predicted by more
          than two gene prediction programs over most of their length are
          annotated as 'hypothetical' proteins. Genes encoding tRNAs are
          predicted by tRNAscan-SE (Sean Eddy,
          http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
          identified by repeatmasker (Arian Smit,
          http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
          genomic sequence that are not annotated as genes but have predicted
          exons by GRAIL are annotated as misc features.
FEATURES          Location/Qualifiers
     source          1..86017
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="GPA1"
                     /clone="T22F11"
     misc_feature    complement(94..173)
                     /note="exon predicted by xgrail, quality marginal"
     repeat_region   complement(384..444)
                     /rpt_family="(TA)n"
     repeat_region   complement(408..532)
                     /rpt_family="(TAA)n"
     repeat_region   complement(558..711)
                     /rpt_family="AT_rich"
     misc_feature    complement(715..778)
                     /note="hypothetical protein"
                     /codon_start=1
     mRNA            complement(join(<971..1026,1124..1295,1815..1853,2757..2867,
                     3920..>4225)
     gene            <971..>4225
                     /gene="T22F11.1"
     CDS             join(971..1026,1124..1295,1815..1853,2757..2867,
                     3920..4225)
                     /gene="T22F11.1"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD23652.1"
                     /db_xref="GI:4567238"
                     /translation="MQLEPRINTFSMEAMFTFSSTFLYDSIVDSSRPILATSLLGCDV

     TDPRRATTACCICEETLSELHMTRPTQVKVAGMSCRRQTKKMEQNDYIKNENDHAI
     KLQDFGYLGPERNEGGIFNRRKLDVKKNTWTKRIYVLPFMWNNIVATTKRIVGMIG
     TSEIVLSPYDLSDPYLFYYNLETNNVREVGIQGLGAFETSTVVHLFVDYEEDVKLYE
     CLKIWLEIS"
     repeat_region   2309..2358
                     /rpt_family="(GAA)n"
     tRNA            complement(2929..3001)
                     /gene="T22F11.2"
                     /product="tRNA-Ala"
                     /anticodon=(pos:2966..2968,aa:Ala)
     gene            complement(2929..3001)
                     /gene="T22F11.2"
     repeat_region   3252..3353
                     /rpt_family="AT_rich"
     repeat_region   4680..4760
                     /rpt_family="(CAT)n"
     mRNA            complement(join(<4693..4846,4930..5142,5399..5666,
                     5675..5699,6086..>6478))
                     /gene="T22F11.3"
     gene            complement(<4693..>6478)
                     /gene="T22F11.3"
                     /note="predicted by genscan;contains cysteine rich domain"
     CDS             complement(join(4693..4846,4930..5142,5399..5666,
                     5675..5699,6086..6478))
                     /gene="T22F11.3"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD23653.1"
                     /db_xref="GI:4567239"
                     /translation="MERDDINLIFQKRRIDSATTENSNPKGDDSKVVNSASPATQS
     YRLYFKGLYSDQILAEGGKRIVNAGFGVAICNTDEFLFEIKESLSNAEISRGVEIV
     ALIRGLSECLSFGFMNVVYCDDHQIYQYELFSSQIIERSLRRKLLVEEVKRLREQMT
     FSEAVLVARNDVKFAYRLAREEIVSKSSSVNVKEAQGETCVICLEETVADRMFFTDKC
     LHRYCFSCVKQTCVKCNGLFCIDCKVPSHSDLSCADYKKLHPELLVDDIKLKLLANEN
     MWRQCVMCRHLIELSDGCNHMTCRCGYQFCYGCGIEWKKNQDTCPSGCREMGHGDADS
     DDDDDVDDDEDDYCM"
     repeat_region   complement(6558..6692)
                     /rpt_family="AT_rich"
     repeat_region   6849..6914
                     /rpt_family="AT_rich"
     mRNA            complement(join(<7046..7232,7601..8334,8950..9450,
                     9922..>10140))
                     /gene="T22F11.4"
     gene            complement(<7046..>10140)
                     /gene="T22F11.4"
                     /note="predicted by genscan"
     CDS             complement(join(7046..7232,7601..8334,8950..9450,
                     9922..10140))
                     /gene="T22F11.4"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD23654.1"
                     /db_xref="GI:4567240"
                     /translation="MERHDQELRYQKRRLRPCSFPCRWIQRQRPATSDSNRQQPATA
     TSDNIHCSFVCRWTQRPTTATIANNCRCVRLIAKNFSDQKLRSLIAGDYRFWSLNQR
     QANEQLLVVAGFSRQRANEQGLNSIITGDIPNSKGEGSPAKFMDNVVYRLCFKG
     LVNYEAVADDVEKAVKAGFGVAICDEKDNLLHEIKESLRDIEISRRGVEIMALVRGLS
     ESFDLGMRNVVIFCDDDWIYQSIIGRGKSKKKIDHLVEEVQGILEKMACIDAVLVARN
     DVKFAFRLAREAIGRNSVDVNAEQGETCGIFFEETDVEHMFVTEKCLHRHCFPCVKQH
     VRVKLRSGTEPTCLEYGCKFLTLERCSKVLTLKLIEMWKQKMKEDSIPAAERIYCPY
     PNCSMLMSKTELSSEADLSNVRFCVKCCGLFCIDCKVPSHTDLSYDDYKKLHPDPLVD
     DLKLKSLANDKMVRQCVCRHMIELSHGCNHMTCSYDAPRDDQDVVYKYSPEQFFSE
     ERNADGDDYIWDDYNNYGGLTGDKGGFDEGFFRDYSL"
     repeat_region   7348..7461
                     /rpt_family="(CAT)n"
     repeat_region   complement(8674..8705)
                     /rpt_family="AT_rich"
     repeat_region   9519..9594
                     /rpt_family="(TAAA)n"
     repeat_region   9523..9654
                     /rpt_family="(TAA)n"
     repeat_region   9707..9793
                     /rpt_family="(TAAA)n"
```

```
repeat_region    complement(9820..9891)
                 /rpt_family="AT_rich"
repeat_region    10479..10576
                 /rpt_family="(TAAAA)n"
mRNA             complement(join(<11079. .11215,11357. .11526,11611. .12335,
                 12502. .>12591))
                 /gene="T22F11.5"
gene             complement(<11079. .>12591)
                 /gene="T22F11.5"
                 /note="predicted by genscan;contains Zn figer domain (RING
                 finger)"
CDS              complement(join(11079. .11215,11357. .11526,11611. .12335,
                 12502..12591))
                 /gene="T22F11.5"
                 /note="hypothetical protein"
                 /codon_start=1
                 /protein_id="AAD23655.1"
                 /db_xref="GI:4567241"
                 /translation="MALIRGLRESFDLGIRQVVVCDDQLLYQYIVGRGKIKKKVVHL
                 VDEFQLLEEMTYTDADLIALNDVNFAFKLAREAIVSRDDVKAEICSICFEETEGERM
                 FFTTEKCVHRHCFPCVKQYVEVKLLSGTVPTCLDDGCKFKLTLESCSKVLTLELIEMW
                 KQKMEDSIPAAERIYCPYPNCSMLMSKTELSSESDLSNDRSCVKCCGLFCIDCKVPS
                 HSDLSCAEYKLHHDPLVDELKLKSLAKDKKWRQCMCRHMIELSHACNHMTCRCGYQ
                 FCYQCEVEWKNDQKTCSSGCLLTGHGYYDYDYNEPEYDFEVDTCNYYSDEEAMTRRE
                 MIRMWNTEQFSEEENADGNDDNWDDYNNYGGLTVFWRL"
misc_feature     12776..12840
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
mRNA             complement(join(<13249. .13371,13455. .13545,13635. .13731,
                 14210. .14306,14651. .14707,16298. .16483,16607. .16849,
                 16970. .17134,17252. .18012,18142. .18250,18718. .18798,
                 19250. .19923,20863. .>20965))
                 /gene="T22F11.6"
gene             complement(<13249. .>20965)

alignment_scores:
         Quality:  40.00                Length:         12
           Ratio:   3.636                 Gaps:          0
Percent Similarity:  91.667     Percent Identity:   58.333

alignment_block:
US-08-653-294-20 x ATAC007070/rev
Align seg 1/1   to reverse of: ATAC007070   from: 1   to: 86017

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeu 12
      |||||:::::|||||||||||  ||||::::::|||
23701 TATCGCTATCGGCATTAGGATCATTAGAGTAATACTC 23666

seq_name: gb_htg3:AC009548

seq_documentation_block:
LOCUS      AC009548   99657 bp    DNA             HTG       03-SEP-1999
DEFINITION Homo sapiens chromosome 11 clone 358_I_19 map 11, LOW-PASS SEQUENCE
           SAMPLING.
ACCESSION  AC009548
VERSION    AC009548.2   GI:5822701
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 99657)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 11, clone 358_I_19
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 99657)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
           Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
           Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
           Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
           Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
           Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
```

```
           Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
           Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
           Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
           Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
           Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
           Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
           Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
           Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
           Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Sep 3, 1999 this sequence version replaced gi:5788096.
           All repeats were identified using RepeatMasker: Smit,A.F.A. &
           Green,P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html.
         * NOTE: This record contains 113 individual
         * sequencing reads that have not been assembled into
         * contigs. Runs of N are used to separate the reads
         * and the order in which they appear is completely
         * arbitrary. Low-pass sequence sampling is useful for
         * identifying clones that may be gene-rich and allows
         * overlap relationships among clones to be deduced.
         * However, it should not be assumed that this clone
         * will be sequenced to completion. In the event that
         * the record is updated, the accession number will
         * be preserved.
         *        1    802:   contig of 802 bp in length
         *              gap of unknown length
         *      803   1624:   contig of 822 bp in length
         *              gap of unknown length
         *     1625   2425:   contig of 801 bp in length
         *              gap of unknown length
         *     2426   3241:   contig of 816 bp in length
         *              gap of unknown length
         *     3242   4054:   contig of 813 bp in length
         *              gap of unknown length
         *     4055   4865:   contig of 811 bp in length
         *              gap of unknown length
         *     4866   5672:   contig of 807 bp in length
         *              gap of unknown length
         *     5673   6502:   contig of 830 bp in length
         *              gap of unknown length
         *     6503   7319:   contig of 817 bp in length
         *              gap of unknown length
         *     7320   8127:   contig of 808 bp in length
         *              gap of unknown length
         *     8128   8932:   contig of 805 bp in length
         *              gap of unknown length
         *     8933   9862:   contig of 930 bp in length
         *              gap of unknown length
         *     9863  10755:   contig of 893 bp in length
         *              gap of unknown length
         *    10756  11571:   contig of 816 bp in length
         *              gap of unknown length
         *    11572  12387:   contig of 816 bp in length
         *              gap of unknown length
         *    12388  13196:   contig of 809 bp in length
         *              gap of unknown length
         *    13197  14011:   contig of 815 bp in length
         *              gap of unknown length
         *    14012  14828:   contig of 817 bp in length
         *              gap of unknown length
         *    14829  15639:   contig of 811 bp in length
         *              gap of unknown length
         *    15640  16504:   contig of 865 bp in length
         *              gap of unknown length
         *    16505  17324:   contig of 820 bp in length
         *              gap of unknown length
         *    17325  18130:   contig of 806 bp in length
         *              gap of unknown length
         *    18131  18937:   contig of 807 bp in length
         *              gap of unknown length
```

```
*  18938  19792: contig of 855 bp in length
*                gap of unknown length
*  19793  20604: contig of 812 bp in length
*                gap of unknown length
*  20605  21420: contig of 816 bp in length
*                gap of unknown length
*  21421  22242: contig of 822 bp in length
*                gap of unknown length
*  22243  23048: contig of 806 bp in length
*                gap of unknown length
*  23049  23890: contig of 842 bp in length
*                gap of unknown length
*  23891  24707: contig of 817 bp in length
*                gap of unknown length
*  24708  25523: contig of 816 bp in length
*                gap of unknown length
*  25524  26336: contig of 813 bp in length
*                gap of unknown length
*  26337  27148: contig of 812 bp in length
*                gap of unknown length
*  27149  27956: contig of 808 bp in length
*                gap of unknown length
*  27957  28761: contig of 805 bp in length
*                gap of unknown length
*  28762  29572: contig of 811 bp in length
*                gap of unknown length
*  29573  30373: contig of 801 bp in length
*                gap of unknown length
*  30374  31177: contig of 804 bp in length
*                gap of unknown length
*  31178  31982: contig of 805 bp in length
*                gap of unknown length
*  31983  32791: contig of 809 bp in length
*                gap of unknown length
*  32792  33626: contig of 835 bp in length
*                gap of unknown length
*  33627  34496: contig of 870 bp in length
*                gap of unknown length
*  34497  35333: contig of 837 bp in length
*                gap of unknown length
*  35334  36800: contig of 1467 bp in length
*                gap of unknown length
*  36801  37601: contig of 801 bp in length
*                gap of unknown length
*  37602  38406: contig of 805 bp in length
*                gap of unknown length
*  38407  39214: contig of 808 bp in length
*                gap of unknown length
*  39215  40038: contig of 824 bp in length
*                gap of unknown length
*  40039  40945: contig of 907 bp in length
*                gap of unknown length
*  40946  41921: contig of 976 bp in length
*                gap of unknown length
*  41922  42857: contig of 936 bp in length
*                gap of unknown length
*  42858  43829: contig of 972 bp in length
*                gap of unknown length
*  43830  44705: contig of 876 bp in length
*                gap of unknown length
*  44706  45648: contig of 943 bp in length
*                gap of unknown length
*  45649  46524: contig of 876 bp in length
*                gap of unknown length
*  46525  47465: contig of 941 bp in length
*                gap of unknown length
*  47466  48319: contig of 854 bp in length
*                gap of unknown length
*  48320  49249: contig of 930 bp in length
*                gap of unknown length
*  49250  50175: contig of 926 bp in length
*                gap of unknown length
*  50176  51105: contig of 930 bp in length
*                gap of unknown length

*  51106  52091: contig of 986 bp in length
*                gap of unknown length
*  52092  53040: contig of 949 bp in length
*                gap of unknown length
*  53041  53966: contig of 926 bp in length
*                gap of unknown length
*  53967  54835: contig of 869 bp in length
*                gap of unknown length
*  54836  55789: contig of 954 bp in length
*                gap of unknown length
*  55790  56674: contig of 885 bp in length
*                gap of unknown length
*  56675  57631: contig of 957 bp in length
*                gap of unknown length
*  57632  58514: contig of 883 bp in length
*                gap of unknown length
*  58515  59397: contig of 883 bp in length
*                gap of unknown length
*  59398  60274: contig of 877 bp in length
*                gap of unknown length
*  60275  61163: contig of 889 bp in length
*                gap of unknown length
*  61164  62071: contig of 908 bp in length
*                gap of unknown length
*  62072  62952: contig of 881 bp in length
*                gap of unknown length
*  62953  63827: contig of 875 bp in length
*                gap of unknown length
*  63828  64798: contig of 971 bp in length
*                gap of unknown length
*  64799  65764: contig of 966 bp in length
*                gap of unknown length
```

```
alignment_scores:
          Quality:   40.00            Length:         13
            Ratio:    4.000              Gaps:          0
Percent Similarity:   76.923    Percent Identity:  61.538

alignment_block:
US-08-653-294-20 x AC009548

Align seg 1/1  to: AC009548   from: 1  to: 99657

        1 TyrArgLeuAlaIleAargIle***ArgIleLeuLeuArg 13
          |||||||||||||| |||  :::::::::||||||
    96875 TACCGCCTTGCGCAAGGGCAAGGGCAGACCGAGTAGTTGCGC 96913
```

THIS PAGE BLANK (USPTO)

QM of: US-08-653-294-20 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000   1:28 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+p2n.model -DEV=xlp
 -Q=/cgnl_1/USPTO_spool/US08653294/runat_04022000_160701_15807/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-08-653-294-20
Query length: 14
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 590.520000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:T84234	+	40.00	115.48	78.45	3110	! DNA encoding an autolysin and
N_Geneseq_36:V53557	+	40.00	115.48	78.45	3110	! DNA encoding 2 Staphylococcus
N_Geneseq_36:V74331	+	40.00	104.87	305.85	10470	! Staphylococcus aureus contig
N_Geneseq_36:V69717	+	39.00	109.69	164.86	4031	! Tumour rejection antigen precu
N_Geneseq_36:V69720	+	39.00	109.28	173.78	4225	! Tumour rejection antigen precu
N_Geneseq_36:T48660	+	37.00	135.04	6.39	99	! DNA encoding MAb anti-HBSAg bind
N_Geneseq_36:X24237	-	37.00	121.63	35.64	459	! WO916900 Seq ID 7. Identifying
N_Geneseq_36:X24253	-	37.00	121.63	35.64	459	! WO916894 Seq ID 7. Identifying
N_Geneseq_36:T33641	-	37.00	121.53	72.89	869	! Aspergillus arabinofuranosidase
N_Geneseq_36:Q98553	-	37.00	116.06	106.12	1215	! E.coli L-carnatine dehydratase
N_Geneseq_36:T51339	+	37.00	113.13	162.40	1776	! Coding sequence for alkaline l
N_Geneseq_36:N60309	-	37.00	109.81	196.90	2109	! Equine IFN-omega-1 from pAH61.
N_Geneseq_36:T33646	-	37.00	108.31	244.14	2555	! Aspergillus arabinofuranosidas
N_Geneseq_36:V29571	+	37.00	106.63	496.71	4815	! L.lactis soluble part (F1) ex
N_Geneseq_36:V52237	+	37.00	101.09	723.54	6735	! Streptococcus pneumoniae genom
N_Geneseq_36:X20568	-	37.00	98.16	862.03	7874	! Polynucleotide sequence from t
N_Geneseq_36:X13331	+	37.00	96.79	1.7e+03	14141	! Enterococcus faecalis genome
N_Geneseq_36:X13238	+	37.00	91.68	2.1e-03	17764	! Enterococcus faecalis genome
N_Geneseq_36:T61973	+	36.00	89.68	346.44	2333	! Pseudomonas pseudoalcaligenes
N_Geneseq_36:X20517	-	36.00	103.90	1.1e+03	6761	! Polynucleotide sequence from t
N_Geneseq_36:T28774	+	36.00	94.60	2.6e+03	14042	! Human placental calcium senso
N_Geneseq_36:V05999	+	36.00	88.22	2.6e+03	14042	! Human placenta calcium sensor
N_Geneseq_36:V05995	+	36.00	88.22	2.6e+03	14044	! Human parathyroid calcium sen
N_Geneseq_36:V28776	+	36.00	88.21	2.6e+03	14044	! Human parathyroid calcium sen
N_Geneseq_36:V05997	+	36.00	88.21	2.6e+03	14080	! Human kidney calcium sensor p
N_Geneseq_36:V28775	+	36.00	88.19	2.6e+03	14080	! Human kidney calcium sensor p
N_Geneseq_36:V05996	+	36.00	88.19	2.6e+03	14086	! Human calcium sensor protein
N_Geneseq_36:V28773	+	36.00	88.19	2.6e+03	14086	! Human calcium sensor protein
N_Geneseq_36:V05994	+	36.00	88.19	2.6e+03	14086	! Human calcium protein sensor
N_Geneseq_36:T07264	+	35.00	108.70	187.15	900	! Mutant LAG1 coding sequence. Eu
N_Geneseq_36:T43690	+	35.00	103.65	357.68	1604	! Alcohol acetyltransferase codi
N_Geneseq_36:T15232	+	35.00	103.18	380.25	1694	! Tumor necrosis factor receptor
N_Geneseq_36:T94635	+	35.00	103.18	380.25	1694	! TNF-RI-DD ligand protein clone
N_Geneseq_36:T07263	+	35.00	101.16	492.57	2134	! LAG1 coding sequence. Eukaryot
N_Geneseq_36:T94639	+	35.00	99.87	581.08	2473	! TNF-RI-DD ligand protein clone
N_Geneseq_36:V23687	-	35.00	97.05	834.34	3415	! L.lactis DB1341 pfl gene. Lact
N_Geneseq_36:Q62924	-	35.00	88.66	2.4e+03	8920	! Carbamoyl-phosphate-synthetase
N_Geneseq_36:X20546	+	35.00	83.21	4.9e+03	16636	! Polynucleotide sequence from
N_Geneseq_36:X20569	-	35.00	81.31	6.3e+03	20682	! Polynucleotide sequence from
N_Geneseq_36:T80649	-	34.00	118.82	51.14	189	! Type II topoisomerase database
N_Geneseq_36:X30892	+	34.00	108.83	184.22	593	! Streptococcus pneumoniae genomi
N_Geneseq_36:T96357	-	34.00	107.89	207.71	660	! CD28 cDNA. New xanthene derivat

N_Geneseq_36:X20213	-	34.00	106.68	242.58	758	! Enterococcus faecalis EF110
N_Geneseq_36:Q28837	-	34.00	106.63	244.01	762	! Sequence of the CD28 gene.S
N_Geneseq_36:X19109	+	34.00	106.20	258.05	801	! Rhodosporidium toruloides D-
N_Geneseq_36:X20212	-	34.00	105.30	289.67	888	! Enterococcus faecalis gene E

seq_name: N_Geneseq_36:T84234

seq_documentation_block:
ID T84234 standard; DNA; 3110 BP.
AC T84234;
DT 01-SEP-1998 (first entry)
DE DNA encoding an autolysin and 3 unknown proteins.
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcus gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome; ss.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT CDS   679..1272
FT   /*tag= a
FT CDS   1668..1844
FT   /*tag= b
FT CDS   1885..2064
FT   /*tag= c
FT CDS   2326..2832
FT   /*tag= d
FT   /product= "autolysin"
PN WO9730070-A1.
PD 21-AUG-1997.
PF 19-FEB-1997; US02318.
PR 20-FEB-1996; US-011888.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR WPI; 97-424969/39.
DR P-PSDB; W28340, W28341, W28342, W28343.
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection.
PS Claim 9; Page 977-978; 989pp; English.
CC The present sequence encodes 3 Staphylococcus aureus proteins of
CC unknown function and a protein, that, based on homology with a
CC known Staphylococcus aureus protein, is believed to be an autolysin
CC (EC 3.5.1.28) (N-acetylmuramoyl-L-alanine amidase).
CC The present sequence was obtained from a
CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of Staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The encoded protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
SQ sequence 3110 BP; 1111 A; 488 C;   675 G;   833 T;

alignment_scores:
        Quality: 40.00   Length: 14
        Ratio: 3.333   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 35.714

alignment_block:
US-08-653-294-20 x T84234

Align seg 1/1 to: T84234   from: 1 to: 3110

        1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
          ||:::::||||::::::::::: ||::::||||:::::||||
      785 TACCAGGTAGCCTTGAAAGTTATGAGGGTTATTTATCGGTAT 826

seq_name: N_Geneseq_36:V53557

```
seq_documentation_block:
ID   V53557 standard; DNA: 3110 BP.
AC   V53557;
DT   30-OCT-1998 (first entry)
DE   DNA encoding 2 Staphylococcus aureus proteins of unknown function.
KW   Staphylococcus aureus protein; immune response induction; eye infection;
KW   antibody production; T-cell immune response; gastrointestinal infection;
KW   respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW   central nervous system; kidney infection; urinary tract infection;
KW   antimicrobial compound identification; broad spectrum antibiotic;
KW   therapy; ss.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   CDS             1208..1300
FT                   /tag= a
FT                   /note= "encoded protein shown in W77771"
FT   CDS             1282..1668
FT                   /tag= b
FT                   /note= "encoded protein shown in W77772"
PN   EP-841394-A2.
PD   13-MAY-1998.
PF   24-SEP-1997; 307485.
PR   24-SEP-1996; US-027032.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Burnham MKR, Hodgson JE, Knowles DJC,
PI   Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
PI   Ward JM;
DR   WPI: 98-252940/23.
DR   P-PSDB; W77771, W77772.
PT   New nucleic acid sequences from Staphylococcus aureus WCHU29 -
PT   useful in vaccines and for treatment of bacterial infections of e.g.
PT   respiratory tract and central nervous system
PS   Claim 1; Page 214-216; 390pp; English.
CC   This sequence encodes 2 Staphylococcus aureus proteins of unknown
CC   function, and represents a DNA Sequence of the invention.
CC   The DNA sequences were isolated from Staphylococcus aureus WCHU29
CC   (NCIMB 40771). Host cells containing the DNA sequences are used to
CC   produce polypeptides or fragments. The proteins are used in the treatment
CC   of disease, for inducing an immune response by administering them, to
CC   produce antibody and/or T-cell immune response. Antagonists of the
CC   proteins are used for the inhibition of bacterial polypeptides.
CC   Conditions which may be treated include bacterial infections, especially
CC   respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC   urinary tract, skin, bones and joints. The proteins can also be used to
CC   identify antimicrobial compounds which are broad spectrum antibiotics,
CC   especially useful in the treatment of H. pylori infection.
SQ   Sequence 3110 BP;    488 C;   1111 A;    676 G;     832 T;

alignment_scores:
     Quality:     40.00            Length:        14
     Ratio:       3.333            Gaps:           0
Percent Similarity: 85.714       Percent Identity: 35.714

alignment_block:
US-08-653-294-20 x V53557
Align seg 1/1  to: V53557  from: 1  to: 3110

   1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
     ||:::::||:::::::  |||  ||:::::||       ||||
 785 TACCAGGTAGCGTTGAAAGTTATGAGGGTTATTATCGGTAT 826

seq_name: N_Geneseq_36:V74331

seq_documentation_block:
ID   V74331 standard; DNA: 10470 BP.
AC   V74331;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #20.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
```

```
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scaled skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    361..420
FT                   /tag= a
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    2161..2220
FT                   /tag= b
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    3961..4020
FT                   /tag= c
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    5761..5820
FT                   /tag= d
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    7561..7620
FT                   /tag= e
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
FT   misc_feature    9361..9420
FT                   /tag= f
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI: 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines                                              -
PS   Claim 1; Page 271-277; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scaled skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 10470 BP;   2804 A;   2009 C;   1550 G;    3742 T;
```

alignment_scores:
    Quality: 40.00   Length: 14
    Ratio: 3.333   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 35.714

alignment_block:
US-08-653-294-20 x V74331/rev   ..

Align seg 1/1 to reverse of: V74331 from: 1 to: 10470

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuArgTyr 14
      ||:::::||::::||:::::|| :::::::|| |||||
 9287 TACCAGGTAGCGTTGAAGTTATGAGGGTTATTATCGGTAT 9246

seq_name: N_Geneseq_36:V69717

seq_documentation_block:
ID V69717 standard; cDNA: 4031 BP.
AC V69717;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   257..3475
FT   /*tag= a
PN WO9849184-A1.
PD 05-NOV-1998.
PF 24-APR-1998; U08493.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, De Smet C, Lucas S;
DR WPI; 99-024041/02.
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 26; Page 40-42; 84pp; English.
CC This nucleotide sequence comprises novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 cDNA (see also V69720). MAGE-C1
CC is a novel member of the MAGE family that may be recognised by
CC cytotoxic T cells, leading to lysis of the tumour cells which
CC express it. MACE-C1 and MACE-C2 (see W81546-47) are expressed in a
CC variety of tumours and in normal testis cells, but not by other
CC normal cells. The MAGE-C1 cDNA was isolated from a melanoma
CC LB373-MEL cDNA library using a probe generated from LB-373-MEL
CC RNA by PCR (see V69728-29). It shows homology to MAGE-A1 cDNA (see
CC V69719) and codes for a putative 1072-amino acid protein. The
CC MACE-C1 gene was localised to chromosome region Xq26-q27. MACE-C1
CC and MACE-C2 cDNAs (see V69726) are claimed, as are: expression
CC vectors; transformed or transfected cell lines (e.g. COS and CHO);
CC an isolated TRAP encoded by the cDNAs; a kit useful in a PCR based
CC assay comprising an oligonucleotide having a sequence of
CC nucleotides 18-34 of the 4031 bp MACE-C1 cDNA and an oligonucleotide
CC having a sequence which is complementary to nucleotides 200-217 of
CC the 4031 bp cDNA sequence; a method for determining expression of
CC a MACE-C1 gene using the kit; a polytope comprising a number of
CC tumour rejection antigens derived from MAGE-C1 or MAGE-C2; and a
CC polytope comprising at least one tumour rejection antigen derived
CC from MAGE-C1 or MAGE-C2 and at least one other tumour rejection
CC antigen. MACE-C1 and MACE-C2 can be used in a method for determining
CC the presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
SQ Sequence 4031 BP; 849 A; 1138 C; 877 G; 1167 T;

alignment_scores:
    Quality: 39.00   Length: 13
    Ratio: 3.545   Gaps: 0
Percent Similarity: 84.615   Percent Identity: 46.154

alignment_block:

US-08-653-294-20 x V69717   ..

Align seg 1/1 to: V69717 from: 1 to: 4031

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArg 13
      ||||||::: :::||||:: ::::::::|||||||||
 1317 TACAGAGTTCTCCTGAGAGTGCTCAAAGTGCTTTTGAGG 1355

seq_name: N_Geneseq_36:V69720

seq_documentation_block:
ID V69720 standard; cDNA; 4225 BP.
AC V69720;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   257..3685
FT   /*tag= a
PN WO9849184-A1.
PD 05-NOV-1998.
PF 24-APR-1998; U08493.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, De Smet C, Lucas S;
DR WPI; 99-024041/02.
DR P-PSDB; W81546.
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 1; Page 48-50; 84pp; English.
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 (see W81546). MAGE-C1 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MACE-C1 and MACE-C2 (see W81547) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C1 cDNA was isolated from a melanoma LB373-MEL cDNA by
CC PCR amplification (see V69732-3). It shows homology to MAGE-A1
CC cDNA (see V69719). The MACE-C1 gene was localised to Xq26-q27.
CC MACE-C1 and MACE-C2 cDNAs (see V69726) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MACE-C1
CC gene using the kit; a polytope comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polytope
CC comprising at least one tumour rejection antigen derived from
CC MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC MACE-C1 and MACE-C2 can be used in a method for determining the
CC presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
SQ Sequence 4225 BP; 871 A; 1198 C; 923 G; 1233 T;

alignment_scores:
    Quality: 39.00   Length: 13
    Ratio: 3.545   Gaps: 0
Percent Similarity: 84.615   Percent Identity: 46.154

alignment_block:
US-08-653-294-20 x V69720   ..

, Align seg 1/1 to: V69720 from: 1 to: 4225

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArg 13
      ||||||::: :::||||:: ::::::::|||||||||
 1527 TACAGAGTTCTCCTGAGAGTGCTCAAAGTGCTTTTGAGG 1565

seq_name: N_Geneseq_36:T48660

seq_documentation_block:

ID   T48660 standard; DNA; 99 BP.
AC   T48660;
DT   07-NOV-1997 (first entry)
DE   DNA encoding MAb anti-HBsAg binder sequence, B13, from R26 library.
KW   Functional surrogate: analyte; affinity receptor; immunoreactive group;
KW   mimic; homogenous immunoassay; detection; diagnostic analyte; Chlamydia;
KW   cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;
KW   pregnancy; infectious disease; ferritin; myosin light chain; troponin;
KW   follicle stimulating hormone; human; growth hormone; immunoglobulin E;
KW   prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;
KW   antibody; chorionic gonadotropin; luteinising hormone; cytomegalovirus;
KW   Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;
KW   carcinoembryonic antigen; alpha-fetoprotein; herpes virus; CA125; ss.
OS   Synthetic.
FH   Key   Location/Qualifiers
FT   cds   1..99
FT   /*tag= a
FT   /transl_except= pos:61..63, aa:Arg
FT   /transl_except= pos:70..72, aa:Cys
FT   /note= "No stop codon given"
PN   WO9641172-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U10498.
PR   07-JUN-1995; US-476375.
PA   (CYTO-) CYTOGEN CORP.
PI   Carter JM, Lee-Own FV;
DR   WPI; 97-077284/07.
DR   P-PSDB; W10864.
PT   Labelled functional surrogate of an analyte - useful as competitor
PT   molecule in affinity assays, esp. for detecting large macromolecules
PT   such as ferritin
PS   Claim 51; Page 98; 156pp; English.
CC   This sequence encodes a monoclonal anti-hepatitis B antigen binder
CC   sequence from the R26 library which may be used in affinity assays
CC   invention. The novel labelled conjugate comprises at least one label
CC   attached to a functional surrogate of an analyte of interest. The
CC   surrogate is capable of competing effectively with the analyte for a
CC   limiting amount of an affinity receptor for the analyte. The conjugate
CC   exhibits an activity that is altered upon interaction with the affinity
CC   receptor and this activity can be measured and related to the amount of
CC   the analyte present in a sample. Functional surrogates such as this have
CC   an immunoreactive group that allows the surrogate to compete effectively
CC   and with the analyte for a limiting amount of its affinity receptor.
CC   Functional surrogates are able to mimic naturally occurring analytes.
CC   They can be labelled for use in standard competitive affinity assays
CC   (esp. homogenous immunoassays) for detecting large macromolecules such
CC   as polypeptides, polysaccharides, polynucleotides, glycoproteins and
CC   lipid-containing macromolecules, as well as small haptens. Typical
CC   diagnostic analytes for detection include cardiac or tumour markers,
CC   allergens, hormones related to fertility-pregnancy or analytes associated
CC   with infectious disease. In particular, the assays are useful for
CC   detecting ferritin, follicle stimulating hormone, human growth hormone,
CC   immunoglobulin E, prolactin, parathyroid hormone, human placental
CC   lactogen, hepatitis antigens or antibodies against them, human
CC   chorionic gonadotropin, human luteinising hormone, cytomegalovirus,
CC   Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,
CC   myoglobin, myosin light chain, troponin, carcinoembryonic antigen,
CC   alpha-fetoprotein, prostate-specific antigen and CA125 (a tumour marker).
SQ   Sequence 99 BP;   14 A;   36 C;   28 G;   21 T;

alignment_scores:
        Quality: 37.00   Length: 12
        Ratio: 3.364   Gaps: 0
Percent Similarity: 91.667   Percent Identity: 66.667

alignment_block:
US-08-653-294-20 x T48660/rev

Align seg 1/1 to reverse of: T48660 from: 1 to: 99

  2 ArgLeuAlaIleArgIle***ArgIleLeuLeuArg 13
    ||||||||||::::||::|||||||| ::|||||

96 AGACTGGCCGTTCGAGTGACTCGGATCCAAGTGCGC 61

seq_name: N_Geneseq_36:X24237

seq_documentation_block:
ID   X24237 standard; DNA; 459 BP.
AC   X24237;
DT   01-JUL-1999 (first entry)
DE   WO916900 Seq ID 7.
KW   Gram positive bacteria; covalent coupling; pathogenicity factors
KW   LPXTG-motif dependent C-terminal anchorage; cell surface; antibacterial;
KW   therapeutic agent; extracellular matrix; ss.
OS   Unidentified.
PN   WO916900-A1.
PD   08-APR-1999.
PF   26-SEP-1998; E06136.
PR   29-OCT-1997; EP-118755.
PR   27-SEP-1997; EP-116841.
PA   (EVOT-) EVOTEC BIOSYSTEMS AG.
PI   Goetz F, Pohlner J, Strauss A, Thumm G;
DR   WPI; 99-255104/21.
PT   Identifying nucleic acid that affects adhesion of Gram positive
PT   bacteria
PS   Claim 17; Page 43-44; 58pp; German.
CC   This invention describes a novel method for identifying a nucleic acid
CC   that encodes a polypeptide which affects covalent coupling of
CC   polypeptides to the surface of Gram-positive bacteria. The method
CC   identifies mutations in bacterial factors that (in)directly interact
CC   with the LPXTG-motif dependent C-terminal anchorage of polypeptides,
CC   particularly pathogenicity factors, to the cell surface. The products
CC   of the invention have antibacterial activity and are potentially
CC   useful as therapeutic agents, inhibiting binding of bacteria to
CC   extracellular matrix.
CC   This sequence is represented in the Seq ID listing but does not
CC   appear in the specification.
SQ   Sequence 459 BP;   144 A;   80 C;   62 G;   170 T;

alignment_scores:
        Quality: 37.00   Length: 14
        Ratio: 3.083   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 50.000

alignment_block:
US-08-653-294-20 x X24237/rev

Align seg 1/1 to reverse of: X24237 from: 1 to: 459

    1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
      |||:||:   :::||| :|||||||||:::|||||
  220 TATCAAGTCAGCAAAAAATTGAAAGAATACTTTCAGATAT 179

seq_name: N_Geneseq_36:X24253

seq_documentation_block:
ID   X24253 standard; DNA; 459 BP.
AC   X24253;
DT   01-JUL-1999 (first entry)
DE   WO916894 Seq ID 7.
KW   Active agent; covalent coupling; Gram-positive bacteria; inhibitor;
KW   adhesion; extracellular matrix; LPXTG motif-dependent anchorage;
KW   pathogenicity factor; bacterial cell wall; therapeutic agent; ss.
OS   Unidentified.
PN   WO916894-A1.
PD   08-APR-1999.
PF   26-SEP-1998; E06137.
PR   29-OCT-1997; EP-118756.
PR   27-SEP-1997; EP-116840.
PA   (EVOT-) EVOTEC BIOSYSTEMS AG.
PI   Goetz F, Pohlner J, Strauss A, Thumm G;
DR   WPI; 99-255102/21.
PT   Identifying agents that affect adhesion of Gram positive bacteria
PS   Disclosure; Page 40-41; 51pp; German.

CC   This invention describe a novel method for identifying an active agent
CC   that affects covalent coupling of polypeptides to the surface of
CC   Gram-positive bacteria. The agents identified inhibit
CC   polypeptide-induced adhesion of bacteria to extracellular matrix in the
CC   host. The active agent effects the LPXTG motif-dependent anchorage of
CC   pathogenicity factors to the bacterial cell wall and are potentially
CC   useful as therapeutic agents, inhibiting binding of bacteria to
CC   extracellular matrix in the host.
CC   This sequence is not described in the specification but is represented
CC   in the sequence ID listing.
SQ   Sequence 459 BP; 144 A;   80 C;   62 G;   170 T;

alignment_scores:
         Quality: 37.00   Length: 14
           Ratio: 3.083   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 50.000

alignment_block:
US-08-653-294-20 x X24253/rev   ..

Align seg 1/1 to reverse of: X24253   from: 1 to: 459

     1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
       ||:::::||::: :::||| ||||||||:::::|||||
   220 TATCAAGTCAGCAAAAAATGAAAGAATACTTTTCAGATAT 179

seq_name: N_Geneseq_36:T33641

seq_documentation_block:
ID   T33641 standard; DNA; 869 BP.
AC   T33641;
DT   11-DEC-1996 (first entry)
DE   Aspergillus arabinofuranosidase gene promoter.
KW   Promoter; arabinofuranosidase; AbfC; arabinoxylan;
KW   viscosity modifier; food; feedstuff; ss.
OS   Aspergillus niger strain 3M43.
PN   WO9629416-A1.
PD   26-SEP-1996.
PF   11-MAR-1996; E01009.
PR   17-MAR-1995; GB-005479.
PA   (DANI-) DANISCO AS.
PI   Baruch A, Madrid SM, Rasmussen P;
DR   WPI; 96-443191/44.
PT   Aspergillus arabinofuranosidase - useful for degradation of
PT   arabinoxylan
PS   Claim 8; Page 48; 105pp; English.
CC   The promoter (T33641) of the arabinofuranosidase gene of Aspergillus
CC   niger 3M43 was identified in a full-length gene (T33646) isolated
CC   from a genomic DNA library. The promoter is strongly repressed by
CC   glucose and induced by intermediates of xylose metabolism. It
CC   can be used to control the expression of a gene of interest, e.g.
CC   A. niger arabinofuranosidase or a heterologous gene, in a host
CC   cell, pref. Aspergillus or a transgenic plant.
SQ   Sequence 869 BP; 224 A;   209 C;   193 G;   243 T;

alignment_scores:
         Quality: 37.00   Length: 12
           Ratio: 3.700   Gaps: 0
Percent Similarity: 83.333   Percent Identity: 58.333

alignment_block:
US-08-653-294-20 x T33641   ..

Align seg 1/1 to: T33641   from: 1 to: 869

     1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeu 12
       ||::::||| :::::||: ||||||||||||||
   165 TACCAACTTTTAGTTCGATTGATCAGAATACTTTTG 200

seq_name: N_Geneseq_36:Q98553

---

seq_documentation_block:
ID   Q98553 standard; DNA; 1215 BP.
AC   Q98553;
DT   28-FEB-1996 (first entry)
DE   E.coli L-carnitine dehydratase gene caiB.
KW   E.coli; L-carnitine dehydratase; caiB; carnatine operon; CaiE; rat;
KW   intestine; chromosomal DNA library; probe; expression plasmid; Bacterium;
KW   crotonobetaine; long chain fatty acid; transport; mitochondrial membrane;
KW   haemodialysis; cardiac; ds.
OS   Escherischia coli.
PN   FR2715167-A1.
PD   21-JUL-1995.
PF   20-JAN-1994; 000807.
PR   20-JAN-1994; FR-000807.
PA   (NASC-) INST NAT SCI APPLIQUES LYON.
PI   Eichler K, Kleber H, Mandrand-Berthelot M;
DR   WPI; 95-256580/34.
DR   P-PSDB; R80277.
PT   DNA fragments encoding E.coli L-carnitine dehydratase and CaiE
PT   protein - and related plasmids, transformed E.coli and recombinant
PT   proteins, useful in enantioselective prodn of L-carnitine
PS   Claim 1; Page 17-18; 32pp; French.
CC   The nucleotide sequence of the E.coli L-carnitine dehydratase (CDH) gene
CC   designated caiB. The sequence forms part of the carnatine operon which
CC   includes the accessory protein CaiE encoded by the caiE gene (Q98554).
CC   The sequences were obtained from E.coli strain 044K74 which was isolated
CC   from rat intestine. The caiB gene was obtained by screening an E.coli
CC   chromosomal DNA library with the probes Q98555-7, based on the N-terminal
CC   sequence derived from the purified enzyme. The 1.3 kb MluI-MluI caiB
CC   gene fragment was subcloned into the expression plasmid pT7-6 for
CC   production of the enzyme in E.coli. Bacteria expressing the L-carnatine
CC   dehydratase can be used in the microbiological prodn. of L-carnatine from
CC   crotonobetaine. L-carnatine is an essential factor in long chain fatty
CC   acid transport across internal mitochondrial membranes. It can be used
CC   to treat L-carnatine-deficiency syndromes, haemodialysis and cardiac
CC   patients and to stimulate prodn. of monoclonal antibodies.
SQ   Sequence 1215 BP; 313 A;   330 C;   311 G;   261 T;

alignment_scores:
         Quality: 37.00   Length: 13
           Ratio: 3.700   Gaps: 0
Percent Similarity: 76.923   Percent Identity: 76.923

alignment_block:
US-08-653-294-20 x Q98553   ..

Align seg 1/1 to: Q98553   from: 1 to: 1215

     2 ArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
       ||: ||| :|||||||||| ||||||||||||||
    27 CGGACCGCTGGCCGGATGCCGGTTGCTTCTCCGGTAT 65

seq_name: N_Geneseq_36:T51339

seq_documentation_block:
ID   T51339 standard; DNA; 1776 BP.
AC   T51339;
DT   11-NOV-1997 (first entry)
DE   Alkaline liquefying alpha-amylase.
KW   Coding sequence for alkaline liquefying alpha-amylase.
KW   starch-related polysaccharide; Bacillus; alpha-amylase; detergent;
KW   alpha-1,4-glucosidic link; alkalophilic Bacillus; enzyme; surfactant resistance;
KW   dish-washing detergent; starch; ds.
OS   Bacillus species KSM-AP1378.
FH   Key   Location/Qualifiers
FT   cds   145..1695
FT   /*tag= a
FT   /product= alkaline liquefying alpha-amylase
PN   WO9700324-A1.
PD   03-JAN-1997.
PF   14-JUN-1996; J01641.
PR   14-JUN-1995; JP-147257.

PA (KAOS) KAO CORP.
PI Ara K, Hatada Y, Ito S, Kawai S, Ozaki K;
DR WPI; 97-118708/11.
DR P-PSDB; W11326.
PT DNA encoding alkaline liquefying alpha-amylase - useful in
PT dish-washing and laundry detergents for removal of starch dirts
PS Claim 8; Page 23-26; 40pp; English.
CC This sequence represents the coding sequence for an alkaline liquefying
CC alpha-amylase. Alpha-amylase is an enzyme that acts on starch-related
CC polysaccharides, hydrolysing the alpha-1,4-glucoside bond of the
CC polysaccharide molecule. Alkaline liquefying alpha-amylases exhibit
CC resistance to surfactants used in detergents, and decompose starch or
CC starch-related polysaccharides in a highly random manner. The Bacillus
CC species KSM-AP1378, from which this sequence was isolated, is an
CC alkalophilic Bacillus strain. It was isolated from soil in the vicinity
CC of the city of Tochigi. The enzyme is useful in improving the efficiency
CC of dish-washing and laundry detergents, particularly on starch dirts.
SQ Sequence 1776 BP;   305 C;   575 A;   417 G;   479 T;

alignment_scores:
        Quality: 37.00   Length: 13
        Ratio: 3.700   Gaps: 0
Percent Similarity: 76.923   Percent Identity: 69.231

alignment_block:
US-08-653-294-20 x T51339   ..

Align seg 1/1 to: T51339 from: 1 to: 1776

       2 ArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
         ||||| :::||||||| |||||||||||||
    1320 AGGTTACCCTTCCGTATTTTACGGTGATTACTACGGTAT 1358

seq_name: N_Geneseq_36:N60309

seq_documentation_block:
ID N60309 standard; DNA; 2109 BP.
AC N60309;
DT 10-JUN-1991 (first entry)
DE Equine IFN-omega-1 from pAH61.
KW IFN-omega-1; equine; interferon; ss.
OS Equus cabalus.
FH Key   Location/Qualifiers
FT cds   1421..2005
FT   /tag= a
FT   /product= IFN-omega-1
FT signal_peptide 1421..1490
FT   /tag= b
FT mat_peptide   1491..2005
FT   /tag= c
PN EP-186098-A.
PD 02-JUL-1986.
PF 17-DEC-1985; 116083.
PF 18-DEC-1984; DE-446122.
PA (BOEH) BOEHRINGER INGELHEIM.
PI Himmler A, Hauptmann R, Hauel N, Adolf G, Swetley P;
DR WPI; 86-170649/27.
DR P-PSDB; P60400.
PT New equine and canine interferon - and recombinant DNA molecules
PT coding for them, and transformed cells.
PS Disclosure; Fig 12; 149pp; German.
CC Microorganisms transformed with the recombinant sequence produce IFN
CC which is useful therapeutically in veterinary medicine.
CC See also N60306-13 and N60938.
SQ Sequence 2109 BP;   476 C;   609 A;   471 G;   553 T;

alignment_scores:
        Quality: 37.00   Length: 14
        Ratio: 3.364   Gaps: 0
Percent Similarity: 78.571   Percent Identity: 50.000

---

alignment_block:
US-08-653-294-20 x N60309/rev   ..

Align seg 1/1 to reverse of: N60309 from: 1 to: 2109

       1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
         ||| :::::::::||| |||||||||||||::||
    1896 TATTTCTTCGTGTCAGGTACAGACGGGATTCCTGAAGTAC 1855

seq_name: N_Geneseq_36:T33646

seq_documentation_block:
ID T33646 standard; DNA; 2555 BP.
AC T33646;
DT 11-DEC-1996 (first entry)
DE Aspergillus arabinofuranosidase gene.
KW Arabinofuranosidase; AbfC; arabinoxylan; viscosity modifier; food;
KW feedstuff; ds.
OS Aspergillus niger strain 3M43.
FH Key   Location/Qualifiers
FT promoter   1..869
FT   /tag= a
FT cds   870..1757
FT   /tag= b
FT signal_peptide 870..947
FT   /tag= c
FT mat_peptide   948..1754
FT   /tag= d
FT terminator   1755..2555
FT   /tag= e
PN WO9624916-A1.
PD 26-SEP-1996.
PF 11-MAR-1996; E01009.
PF 17-MAR-1995; GB-005479.
PA (DANI-) DANISCO AS.
PI Baruch A, Madrid SM, Rasmussen P;
DR WPI; 96-443191/44.
DR P-PSDB; W00810.
PT Aspergillus arabinofuranosidase - useful for degradation of
PT arabinoxylan
PS Example; Page 52-55; 105pp; English.
CC The Aspergillus niger 3M43 arabinofuranosidase gene (T33646)
CC codes for the precursor form (W00810) of the arabinofuranosidase
CC enzyme. It was isolated from a genomic DNA library by screening
CC with a PCR clone obtd. by amplification of A. niger DNA using
CC primers (see also T33544-45) based on an isolated peptide (W00806)
CC the enzyme. The mature enzyme coding sequence (see also T33640)
CC can be used for prodn. of arabinofuranosidase (W04167) in
CC transformed hosts, pref. Aspergillus or a transgenic plant. The
CC promoter (see also T33641), terminator (T33642) and signal sequence
CC (T33643) may also be used to control expression of the
CC arabinofuranosidase gene, or other gene of interest, in a host
CC cell.
SQ Sequence 2555 BP;   632 A;   651 C;   615 G;   657 T;

alignment_scores:
        Quality: 37.00   Length: 12
        Ratio: 3.700   Gaps: 0
Percent Similarity: 83.333   Percent Identity: 58.333

alignment_block:
US-08-653-294-20 x T33646   ..

Align seg 1/1 to: T33646 from: 1 to: 2555

       1 TyrArgLeuAlaIleArgIle***ArgIleLeuLeu 12
         ||:|:::||| ::||||::: |||||||||||
     165 TACCAACTTTAGTTCGATTGATTCAGAATACTTTTG 200

seq_name: N_Geneseq_36:V29571

seq_documentation_block:

```
ID   V29571 standard: DNA; 4815 BP.
AC   V29571;
DT   04-AUG-1998  (first entry)
DE   L. lactis soluble part (F1) exhibiting ATPase activity encoding DNA.
KW   Biomass production; uncoupled ATPase; FOF1 ATPase; membrane bound;
KW   F1; Lactococcus lactis; ds.
OS   Lactococcus lactis.
FH   Key             Location/Qualifiers
FT   CDS             26..553
FT                   /*tag= a
FT                   /product= "ATPase subunit"
FT                   /gene= "atpH"
FT                   /note= "delta subunit of the F1 portion of FOF1 ATPase"
FT   CDS             742..2244
FT                   /*tag= b
FT                   /product= "ATPase subunit"
FT                   /gene= "atpA"
FT                   /note= "alpha subunit of the F1 portion of FOF1 ATPase"
FT   CDS             2260..3129
FT                   /*tag= c
FT                   /product= "ATPase subunit"
FT                   /gene= "atpG"
FT                   /note= "gamma subunit of the F1 portion of FOF1 ATPase"
FT   CDS             3301..4710
FT                   /*tag= d
FT                   /product= "ATPase subunit"
FT                   /gene= "atpD"
FT                   /note= "beta subunit of the F1 portion of FOF1 ATPase"
PN   WO9810089-A1.
PD   12-MAR-1998.
PF   08-SEP-1997; DK0373.
PR   06-SEP-1996; DK-000963.
PA   (JENS/) JENSEN P R.
PI   Snoep JL, Westerhoff HV;
DR   WPI: 98-193637/17.
DR   P-PSDB; W56790, W56791, W56792, W56793.
DR   P-PSDB; W56793, W56792, W56793.
PT   Method improving production of biomass or a desired product - by
PT   expressing an uncoupled ATPase activity in the cell
PS   Claim 16; Pages 35-41; 78pp; English.
CC   This DNA encodes the soluble part (F1) of membrane bound (FOF1 type) H+
CC   - ATPase or a portion of F1 exhibiting ATPase activity. The DNA is
CC   derived from Lactococcus lactis subsp. cremoris strain MG1363. This is
CC   used in a novel method for improving the production of biomass or a
CC   desired product from a cell. The method comprises expressing an uncoupled
CC   ATPase activity in the cell to induce conversion of ATP to ADP without
CC   primary effects on other cellular metabolites or functions and incubating
CC   the cell with a suitable substrate to produce the biomass or product. The
CC   expression is directed using a vector including DNA encoding the soluble
CC   part (F1) of the membrane bound (FOF1 type) H+-ATPase or a portion of F1
CC   exhibiting ATPase activity, the DNA being derived from Lactococcus lactis
CC   subsp. cremoris, Lactococcus lactis subsp. Lactis, Streptococcus
CC   thermophilus, Phaffia rhodozyma or Trichoderma reesei, where the DNA is
CC   under the control of a promoter. An ideal ATPase is the membrane bound
CC   H+ ATPase. This enzyme complex consists of two parts, the membrane
CC   integral part and the (FO) and the cytoplasmic part (F1). Together the
CC   two parts couple the hydrolysis of ATP and ADP to the translocation of
CC   protons across the cytoplasmic membrane, or vice versa. The proton
CC   gradient is used to drive ATP synthesis from ADP and Pi. The method can
CC   be used for optimising the formation of biomass or a desired product,
CC   e.g. the product may be lactic acid which results in the acidification of
CC   dairy products. The method is more efficient than currently used methods
CC   of biomass production.
SQ   Sequence 4815 BP; 1491 A; 897 C; 981 G; 1446 T;
```

```
alignment_scores:
            Quality:     37.00            Length:       14
            Ratio:       3.083            Gaps:          0
   Percent Similarity: 85.714    Percent Identity: 42.857

alignment_block:
US-08-653-294-20 x V29571  ..
```

Align seg 1/1 to: V29571 from: 1 to: 4815

```
                    1 TyrArgLeuAlaIleArgIle***ArgIleLeuArgTyr 14
                      ||:::|||::::::::::  ||:::||::||  ||:||
                 4384 TATGAGTTGCAATGGAAGTTCAACGTGTCTTCAACGCTAC 4425
```

seq_name: N_Geneseq_36:V52237

```
seq_documentation_block:
ID   V52237 standard: DNA; 6735 BP.
AC   V52237;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:104.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI: 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 773-777; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence 6735 BP; 2122 A; 1141 C; 1415 G; 2057 T;
```

```
alignment_scores:
            Quality:     37.00            Length:       12
            Ratio:       3.700            Gaps:          0
   Percent Similarity: 83.333    Percent Identity: 66.667

alignment_block:
US-08-653-294-20 x V52237  ..
```

Align seg 1/1 to: V52237 from: 1 to: 6735

```
                    3 LeuAlaIleArgIle***ArgIleLeuLeuArgTyr 14
                      |||::||:::|||   |||||||||||||   |||
                 5344 TTATCAATAAAAATCCAGAGAATATTGCTATCTTAC 5379
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 01:29:43 ; Search time 133.56 seconds
                                   (without alignments)
                                     2.660 Million cell updates/sec

Title: US-08-653-294-21
Perfect score: 70
Sequence: 1 AYRLLIKVIRIVLKY 15

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :   A_Geneseq_36:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	52.9	497	1	W49912	Mouse CLK serine/t
2	36	51.4	141	1	W04183	Cellubrevin-4. Iso
3	36	51.4	1226	1	W13825	Yeast transcriptio
4	35	50.0	302	1	W76777	B. malayi ankyrin
5	35	50.0	302	1	W70609	Ankyrin protein PB
6	35	50.0	1784	1	R77223	Tuberous sclerosis
7	34	48.6	37	1	W76789	D. immitis pD1Ankl
8	34	48.6	37	1	W77378	Lytic peptide with
9	34	48.6	303	1	W76774	D. immitis ankyrin
10	34	48.6	303	1	W70606	Ankyrin protein fr
11	34	48.6	395	1	W88282	Pseudomonas mendoc
12	34	48.6	1594	1	P81184	Sequence encoded b
13	34	48.6	1745	1	W76776	D.immitis ankyrin
14	34	48.6	1745	1	W70608	Full length ankyri
15	33	47.1	20	1	R92907	HLA-B2702 CTL modu
16	33	47.1	20	1	R95428	HLA-B2702 84-75-84
17	33	47.1	20	1	W33778	Immunomodulating d
18	33	47.1	550	1	W81351	Human guanine nucl
19	33	47.1	580	1	W81349	Human guanine nucl
20	33	47.1	878	1	Y00868	S. tuberosum isoam
21	33	47.1	2861	1	W27227	Human TRIO phospho
22	32	45.7	12	1	R95429	HLA-B2702 84-79-84
23	32	45.7	12	1	W33798	Peptide B2702,84-7
24	32	45.7	12	1	W33799	Immunomodulating d
25	32	45.7	35	1	P90056	Human derived pept
26	32	45.7	35	1	R15605	ASP-5 analogue (3)
27	32	45.7	227	1	W22173	S.thermophilus exo
28	32	45.7	227	1	W14074	S.thermophilus exo
29	32	45.7	283	1	W41497	Avian cyclin C pro
30	32	45.7	325	1	W41496	Human cyclin C pro
31	32	45.7	374	1	W63694	Human secreted pro
32	32	45.7	443	1	W84264	Protein encoded by
33	32	45.7	626	1	W41501	Human DP.75, a put
34	32	45.7	2410	1	W19723	Cell cycle checkpo
35	32	45.7	2480	1	W19724	Cell cycle checkpo
36	32	45.7	2644	1	W13152	Human ataxia and r
37	32	45.7	2644	1	W84271	A human ATR protei
38	32	45.7	4472	1	R97246	Virulence gene clu
39	31.5	45.0	57	1	Y02689	Human secreted pro
40	31	44.3	23	1	P81601	Sequence of artifi
41	31	44.3	37	1	W83937	Human secreted pro
42	31	44.3	62	1	P80591	Sequence of artifi
43	31	44.3	101	1	R06693	Feline infectious
44	31	44.3	130	1	W90020	Expressed antigen
45	31	44.3	155	1	W98359	H. pylori GHPO 127

ALIGNMENTS

RESULT   1
W49912
ID   W49912 standard; Protein; 497 AA.
AC   W49912;
DT   20-JUL-1998 (first entry)
DE   Mouse CLK serine/threonine kinase mCLK2.
KW   mCLK2; CLK; serine/threonine kinase; protein kinase; LAMMER kinase;
KW   signal transduction; cancer; contraceptive; mouse; therapy;
KW   diagnosis.
OS   Mus musculus.
FH   Key   Location/Qualifiers
FT   Domain   28..45
FT   /note= "nuclear localisation domain"
FT   Domain   163..479
FT   /note= "catalytic domain"
FT   Peptide   388..393
FT   /note= "LAMMER motif"
PN   WO974872 3-A2.
PD   24-DEC-1997.
PF   17-JUN-1997; IB0946.
PR   19-DEC-1996; US-034286.
PR   17-JUN-1996; US-019629.
PR   09-AUG-1996; US-023485.
PR   13-NOV-1996; US-030860.
PR   15-NOV-1996; US-030964.
PA   (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Aoki N, Chen Z, Kharitonenkov AI, Kim YW, Nayler O,
PI   Ullrich A, Wang HY;
DR   WPI: 98-120302/11.
PT   New phosphatase and kinase enzyme(s) - useful in the diagnosis and
PT   treatment of signal transduction disorders
PS   Claim 11; Fig 4; 138pp; English.
CC   This polypeptide comprises novel mouse CLK serine/threonine kinase
CC   mCLK2, from the CLK serine/threonine kinase family of proteins that
CC   regulate RNA splicing in cells. mCLK2 cDNA was cloned from a mouse
CC   embryo 11.5 p.c. 1ZAP cDNA library. The invention discloses the
CC   discovery of novel protein kinases mCLK2, mCLK3 (see W49912) and
CC   mCLK4 (see W49914) of mol.wt. 59.9, 58.5 and 57.2 kDa,
CC   respectively, as well as other novel proteins (see W49906-10)
CC   involved in cellular signal transduction, and provides vectors,
CC   host cells, purified recombinant proteins, methods for identifying
CC   compounds that activate or inhibit the novel proteins, as well as
CC   methods for the diagnosis and treatment of diseases associated with
CC   the novel proteins. Overexpression of CLK serine/threonine kinases
CC   has been implicated in certain types of cancer. Compounds that
CC   inhibit their catalytic activity or disrupt their interactions
CC   with natural binding partners may act as anti-cancer therapeutics.
CC   mCLK related molecules and compounds may also be useful as male
CC   contraceptives.
SQ   Sequence 497 AA;

Query Match   52.9%;   Score 37;   DB 1;   Length 497;
Best Local Similarity   53.8%;   Pred. No. 29;
Matches   7;   Conservative   4;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   3 RLLIKVIRIVLKY 15

```
          |:|:|:|::|| ||
Db    188 RVLVKIIKNVEKY 200

RESULT 2
W04183
ID   W04183 standard; Protein; 141 AA.
AC   W04183;
DT   29-MAY-1997 (first entry)
DE   Cellubrevin-4.
KW   CB-1; CB-2; CB-3; CB-4; human; cellubrevin; study; regulation;
KW   vesicle trafficking; diagnosis; treatment; infection; leukaemia;
KW   traumatic tissue damage; asthma; arthritis; cancer; lymphoma.
OS   Homo sapiens.
PN   WO9629407-A2.
PD   26-SEP-1996.
PF   22-MAR-1996; U03835.
PR   23-MAR-1995; US-409373.
PA   (INCY-) INCYTE PHARM INC.
PI   Hawkins PR, Murry LE, Seilhamer JJ, Stuart SG;
DR   WPI; 96-443183/44.
DR   N-PSDB; T33717.
PT   Isolated human cellubrevin polynucleotide(s) - useful to develop
PT   prods. for diagnosis and treatment of conditions involving abnormal
PT   membrane trafficking.
PS   Claim 25; Page 32; 59pp; English.
CC   W04180-83 are novel human cellubrevins (CB-1, -2, -3 and -4). CBs
CC   can be used for the study and regulation of vesicle trafficking in
CC   normal, and in acute and chronic disease situations, and for the
CC   diagnosis and treatment of conditions caused by infection, traumatic
CC   tissue damage, hereditary disease, e.g. asthma or arthritis, invasive
CC   cancer, leukaemia and lymphoma or other physiologic/pathologic problems
CC   associated with induced, and otherwise abnormal, membrane trafficking.
CC   In particular, the CB-4 polynucleotide (isolated from a cerebellum
CC   library (CRBLNOT01) can be used in a diagnostic test for conditions or
CC   diseases in which its expression is induced, e.g. cerebellar degenerative
CC   diseases or brain tumours, while CB-4 inhibitors can be used to treat
CC   such conditions or diseases
SQ   Sequence 141 AA;

Query Match           51.4%;    Score 36; DB 1; Length 141;
Best Local Similarity 53.3%;    Pred. No. 12;
Matches    8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy     1 AYRLLIKVIRIVLKY 15
         | || :| |||:||
Db   125 AILLLVIIILIVMKY 139

RESULT 3
W13825
ID   W13825 standard; Protein; 1226 AA.
AC   W13825;
DT   04-JUN-1997 (first entry)
DE   Yeast transcription regulatory factor; suppressor of RNA polymerase B;
KW   Transcription regulatory factor; suppressor of RNA polymerase II;
KW   SRB8; RNA polymerase II; holoenzyme; SWI/SNF.
OS   Saccharomyces cerevisiae.
PN   WO9708301-A1.
PD   28-AUG-1996.
PF   28-AUG-1996; U14192.
PR   31-AUG-1995; US-521872.
PR   11-OCT-1995; US-540804.
PR   26-JAN-1996; US-590399.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI   Chao DM, Koleske AJ, Thompson CM, Young RA;
DR   WPI; 97-179258/16.
DR   N-PSDB; T59908.
PT   Purified RNA polymerase II holo:enzyme - comprises RNA polymerase II
PT   and one or more regulatory proteins, pref. suppressor of RNA
PT   polymerase B proteins or SWI/SNF proteins
PS   Claim 11; Fig 10a-b; 154pp; English.
```

```
CC   Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
CC   SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcription
CC   regulatory factors that act as positive and negative regulators of
CC   RNA polymerase II activity, and are components of the RNA polymerase
CC   II holoenzyme. They were identified using methods designed to
CC   identify transcription factors involved in RNA polymerase II
CC   c-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
CC   CTD activity. Genomic clones (T59904-11) for the SRBs have been obtd.
CC   SRBs can be used to treat diseases resulting from alteration or
CC   deletion of the SRB gene, pref. by gene transfer technology. They
CC   can also be used in in vitro transcription of DNA and to identify
CC   cpds. that modify gene transcription.
SQ   Sequence 1226 AA;

Query Match           51.4%;    Score 36; DB 1; Length 1226;
Best Local Similarity 77.8%;    Pred. No. 1.1e+02;
Matches    7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy     2 YRLLIKVIR 10
         | |||:||
Db   999 YHLLIKIR 1007

RESULT 4
W76777
ID   W76777 standard; Protein; 302 AA.
AC   W76777;
DT   15-JAN-1999 (first entry)
DE   B. malayi ankyrin pBmAnk302 protein.
KW   Ankyrin; helminth; parasite; vaccine; therapy; infection;
KW   passive immunogen; cytotoxic agent.
OS   Dirofilaria immitis.
PN   US5827692-A.
PD   27-OCT-1998.
PF   24-APR-1997; 847429.
PR   24-APR-1997; US-847429.
PA   (HESK- ) HESKA CORP.
PI   Blehm ES, Tang L;
DR   WPI; 98-599392/50.
DR   N-PSDB; V63024.
PT   Nucleic acids encoding ankyrins from helminth parasites - useful for
PT   recombinant production of the proteins for use as vaccines and
PT   treatments against helminth infection
PS   Claim 8; Column 137-140; 84pp; English.
CC   W76769-W76777 represent ankyrin proteins isolated from the helminth
CC   parasites Dirofilaria immitis and Brugia malayi. The nucleic acids and
CC   recombinant products are useful for the recombinant production of the
CC   ankyrin polypeptides. These proteins can then be used as vaccines against
CC   parasitic helminth, e.g. D. immitis or B. malayi. They can also be used
CC   for therapy after infection, and to raise antibodies, also for use in
CC   therapeutics, as passive immunogens, or as therapeutics against helminths
CC   on conjugation to cytotoxic agents. The nucleic acids contained in
CC   viruses, may also be used as viral vaccines, and the nucleic acids
CC   themselves or in vectors may be used as genetic vaccines.
SQ   Sequence 302 AA;

Query Match           50.0%;    Score 35; DB 1; Length 302;
Best Local Similarity 60.0%;    Pred. No. 40;
Matches    6; Conservative 3; Mismatches 3; Indels 1; Gaps 0;

Qy     6 IKVIRIVLKY 15
         ||| :|:|||
Db    53 IKVVELLKY 62

RESULT 5
W70609
ID   W70609 standard; Protein; 302 AA.
AC   W70609;
DT   21-JAN-1999 (first entry)
DE   Ankyrin protein PBmAnk302.
```

```
KW   Ankyrin protein; parasitic helminth; heartworm disease;
KW   elephantiasis; hydrocele.
OS   Brugia malayi.
PN   US5824306-A.
PD   20-OCT-1998.
PF   26-FEB-1998; 031485.
PR   24-APR-1997; US-847429.
PR   26-FEB-1998; US-031485.
PA   (HESK-) HESKA CORP.
PI   Blehm ES, Tang Li;
DR   WPI; 98-593373/50.
DR   N-PSDB; V63315.
PT   Dirofilaria and Brugia ankyrin proteins and antibodies - useful for
PT   protection of animals from disease caused by parasitic helminth
PS   Claim 5; Columns 137-140; 84pp; English.
CC   The present sequence represents a Brugia malayi ankyrin protein.
CC   The ankyrin protein, or anti-ankyrin antibodies, may be used to
CC   protect an animal from disease caused by a parasitic helminth,
CC   especially where the disease is heartworm disease, elephantiasis or
CC   hydrocele.
SQ   Sequence   302 AA;

Query Match             50.0%;  Score 35;  DB 1;  Length 302;
Best Local Similarity   60.0%;  Pred. No. 40;
Matches    6;  Conservative   3;  Mismatches   3;  Indels    0;  Gaps    0;

Qy          6 IKVIRIVLKY 15
              ||| ::|||
Db         53 IKVVELLIKY 62

RESULT   6
R77223
ID   R77223 standard; Protein; 1784 AA.
AC   R77223;
DT   04-DEC-1995 (first entry)
DE   Tuberous sclerosis 2 TSC2 gene product.
KW   Tuberous sclerosis 2; TSC2 gene; gene therapy; tumor.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   modified_site  70
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   peptide        76..84
FT                  /label= Repeat_motif
FT   peptide        99..107
FT                  /label= Repeat_motif
FT   peptide        81..102
FT                  /label= Leucine_zipper
FT   region         171..187
FT                  /label= Membrane-spanning_region
FT   modified_site  132
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  154
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  211
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  311
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  389
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"

FT   modified_site  390
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  428
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   region         459..476
FT                  /label= Membrane_spanning region
FT   modified_site  526
FT                  /label= Phosphorylation
FT                  /note= "protein-kinase C phosphorylation
FT                  site"
FT   region         555..572
FT                  /label= Membrane-spanning_region
FT   modified_site  556
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  609
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  655
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  660
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase phosphorylation
FT                  site"
FT   modified_site  716
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  719
FT                  /label= Phosphorylation
FT                  /note= "potential tyrosine-kinase phosphorylation
FT                  site"
FT   modified_site  757
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  790
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   region         804..821
FT                  /label= Membrane-spanning_region
FT   modified_site  923
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  932
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  939
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
FT   modified_site  946
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  970
FT                  /label= Phosphorylation
FT                  /note= "potential casein-kinase 2 phosphorylation
FT                  site"
FT   modified_site  981
FT                  /label= Phosphorylation
FT                  /note= "potential protein-kinase C phosphorylation
FT                  site"
```

```
FT   modified_site
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   active_site     1014
FT                   /label= Phosphorylation
FT                   /note= "potential casein-kinase 2 phosphorylation
FT                   site"
FT   modified_site   1025
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1036
FT                   /label= Glycosylation
FT                   /note= "potential N-linked glycosylation site"
FT   modified_site   1090
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   2003
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase phosphorylation
FT                   site"
FT   modified_site   2005
FT                   /label= Glycosylation
FT                   /note= "potential N-linked glycosylation site"
FT   modified_site   1250
FT                   /label= Phosphorylation
FT                   /note= "potential tyrosine-kinase phosphorylation
FT                   site"
FT   modified_site   1260
FT                   /label= Phosphorylation
FT                   /note= "potential casein-kinase 2 phosphorylation
FT                   site"
FT   modified_site   1269
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1281
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1317
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1318
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1364
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1370
FT                   /label= Phosphorylation
FT                   /note= "potential protien-kinase C phosphorylation
FT                   site"
FT   modified_site   1410
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1426
FT                   /label= Phosphorylation
FT                   /note= "potential casein-kinase 2 phosphorylation
FT                   site"
FT   modified_site   1439
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1446
FT                   /label= Phosphorylation
```

```
FT                   /note= "potential casein-kinase 2 phosphorylation
FT                   site"
FT   modified_site   1464
FT                   /label= Phosphorylation
FT                   /note= "potential protein-kinase C phosphorylation
FT                   site"
FT   modified_site   1491
FT                   /label= Phosphorylation
FT                   /note= "potential casein-kinase 2 phosphorylation
FT                   site"
FT   modified_site   1499
FT                   /label= Glycosylation
FT                   /note= "potential N-linked glycosylation site"
FT   modified_site   1519
FT                   /label= Phosphorylation

Query Match         50.0%;    Score 35;  DB 1;  Length 1784;
Best Local Similarity   26.7%;    Pred. No. 2.6e+02;
Matches    4;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     1 AYRLLIKVIRIVLKY 15
         :| :::: : |:: ||
Db   335 SYEIVLSITRLIKKY 349


RESULT   7
W76789
ID   W76789 standard; Protein; 33 AA.
AC   W76789;
DT   15-JAN-1999 (first entry)
DE   D. immitis pDiAnk1745 Ankyrin-like repeat protein fragment #12.
KW   Ankyrin; helminth; parasite; vaccine; therapy; infection;
KW   passive immunogen; cytotoxic agent.
OS   Dirofilaria immitis.
PN   US5827692-A.
PD   27-OCT-1998.
PF   24-APR-1997; 847429.
PR   24-APR-1997; US-847429.
PA   (HESK-) HESKA CORP.
PI   Blehm ES, Tang L;
DR   WPI; 98-593992/50.
PT   Nucleic acids encoding ankyrins from helminth parasites - useful for
PT   recombinant production of the proteins for use as vaccines and
PT   treatments against helminth infection
PS   Example 2; Column 35-36; 84pp; English.
CC   W76778-W76802 are ankyrin-like repeat peptide fragments isolated from the
CC   helminth parasite Dirofilaria immitis pDiAnk1075 protein. Such ankyrin
CC   nucleic acids and recombinant products are useful for the recombinant
CC   production of the ankyrin polypeptides. These proteins can then be used
CC   as vaccines against parasitic helminth, e.g. D. immitis or B. malayi.
CC   They can also be used for therapy after infection, and to raise
CC   antibodies, also for use in therapeutics, as passive immunogens, or as
CC   therapeutics against helminths on conjugation to cytotoxic agents. The
CC   nucleic acids contained in viruses, may also be used as viral vaccines,
CC   and the nucleic acids themselves or in vectors may be used as genetic
CC   vaccines.
SQ   Sequence   33 AA;

Query Match         48.6%;    Score 34;  DB 1;  Length 33;
Best Local Similarity   50.0%;    Pred. No. 6.1;
Matches    5;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     6 IKVIRIVLKY 15
         ||:: ::|||
Db    15 IKIVELLKKY 24


RESULT   8
W77378
ID   W77378 standard; peptide; 37 AA.
AC   W77378;
DT   14-DEC-1998 (first entry)
```

DE Lytic peptide with alterable function 3.
KW Biologically active peptide; hormone; drug; toxin;
KW lipid bilayer membrane; microorganism; parasite; virus.
OS Synthetic.
PN WO9841535-A2.
PD 24-SEP-1998.
PF 18-MAR-1998; G00799.
PR 18-MAR-1997; GB-005519.
PA (ANMA-) ANMAT TECHNOLOGY LTD.
PI Ajoula HS, Clarke DJ;
DR WPI; 98-521161/44.
PT New modified peptide(s) - obtained by substitution with an amino
PT acid which is modifiable by a reaction and replacing other amino
PT acids which are not to be modified
PS Claim 7; Page 22; 33pp; English.
CC The peptides W77376-W77390 can be modified by the method of the
CC invention by substituting at least one amino acid of the peptide to
CC provide a peptide having at least one amino acid which is modifiable by a
CC reaction and replacing other amino acids in the peptide with amino acids
CC which are not modifiable by the reaction. The methods can be used for
CC the modification of biologically active peptides such as hormones, drugs,
CC toxins and peptides which act on lipid bilayer membranes. The modified
CC peptides can be used e.g. in the body of an animal or plant or parts in
CC order to affect the structure or integrity or permeability of a foreign
CC body such as a microorganism, parasite or virus present in the body of
CC the animal or plant or within the cells of the body of the animal or
CC plant.
SQ Sequence 37 AA;

Query Match 48.6%; Score 34; DB 1; Length 37;
Best Local Similarity 41.7%; Pred. No. 6.9;
Matches 5; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 3 RLLIKVIRIVLK 14
:||:|:;;:||
Db 4 KLLLKLLKLLLK 15

RESULT 9
W76774
ID W76774 standard; Protein; 303 AA.
AC W76774;
DT 15-JAN-1999 (first entry)
DE D. immitis ankyrin pDIank303 protein.
KW Ankyrin; helminth; parasite; vaccine; therapy; infection;
KW passive immunogen; cytotoxic agent.
OS Dirofilaria immitis.
PN US5827692-A.
PD 27-OCT-1998.
PF 24-APR-1997; 847429.
PR 24-APR-1997; US-847429.
PA (HESK-) HESKA CORP.
PI Blehm ES, Tang L;
DR WPI; 98-593992/50.
DR N-PSDB; V63012.
PT Nucleic acids encoding ankyrins from helminth parasites - useful for
PT recombinant production of the proteins for use as vaccines and
PT treatments against helminth infection
PS Claim 8; Column 79-82; 84pp; English.
CC W76769-W76777 Dirofilaria ankyrin proteins isolated from the helminth
CC parasites Dirofilaria immitis and Brugia malayi. The nucleic acids and
CC recombinant products are useful for the recombinant production of the
CC ankyrin polypeptides. These proteins can then be used as vaccines against
CC parasitic helminth, e.g. D. immitis or B. malayi. They can also be used
CC for therapy after infection, and to raise antibodies, also for use in
CC therapeutics, as passive immunogens, or as therapeutics against helminths
CC on conjugation to cytotoxic agents. The nucleic acids contained in
CC viruses, may also be used as viral vaccines, and the nucleic acids
CC themselves or in vectors may be used as genetic vaccines.
SQ Sequence 303 AA;

Query Match 48.6%; Score 34; DB 1; Length 303;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 5; Conservative 5; Mismatches 4; Indels 1; Gaps 0;

Qy 6 IKVIRIVLKY 15
||:|: ::|||
Db 54 IKIVELLKY 63

RESULT 10
W70606
ID W70606 standard; Protein; 303 AA.
AC W70606;
DT 21-JAN-1999 (first entry)
DE Ankyrin protein fragment PDIank303.
KW Ankyrin protein; parasitic helminth; heartworm disease;
KW elephantiasis; hydrocele.
OS Dirofilaria immitis.
PN US5824306-A.
PD 20-OCT-1998.
PF 26-FEB-1998; 031485.
PR 24-APR-1997; US-847429.
PR 26-FEB-1998; US-031485.
PA (HESK-) HESKA CORP.
PI Blehm ES, Tang L;
DR WPI; 98-593373/50.
DR N-PSDB; V63312.
PT Dirofilaria and Brugia ankyrin proteins and antibodies - useful for
PT protection of animals from disease caused by parasitic helminth
PS Claim 5; Columns 79-82; 84pp; English.
CC The present sequence represents part of a Dirofilaria immitis ankyrin
CC protein. The ankyrin protein, or anti-ankyrin antibodies, may be used
CC to protect an animal from disease caused by a parasitic helminth,
CC especially where the disease is heartworm disease, elephantiasis or
CC hydrocele.
SQ Sequence 303 AA;

Query Match 48.6%; Score 34; DB 1; Length 303;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 5; Conservative 5; Mismatches 4; Indels 1; Gaps 0;

Qy 6 IKVIRIVLKY 15
||:|: ::|||
Db 54 IKIVELLKY 63

RESULT 11
W88282
ID W88282 standard; Protein; 395 AA.
AC W88282;
DT 12-APR-1999 (first entry)
DE Pseudomonas mendocina para-hydroxybenzoate hydroxylase.
KW Para-hydroxybenzoate hydroxylase; para-hydroxybenzoic acid; PHBA;
KW PobA-2.
OS Pseudomonas mendocina.
PN WO9856920-A1.
PD 17-DEC-1998.
PF 11-JUN-1998; U12072.
PR 03-JUN-1998; US-049556.
PR 13-JUN-1997; US-049556.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI Chen KK, Grelak RL;
DR WPI; 99-060332/05.
DR N-PSDB; V84271.
PT Production of para-hydroxybenzoate - using pobA(-) Pseudomonas
PT strains which have a toluene degradation pathway and do not produce
PT para-hydroxybenzoate hydroxylase
PS Claim 4; Page 39-41; 60pp; English.
CC This is the amino acid sequence of a Pseudomonas mendocina
CC para-hydroxybenzoate hydroxylase (PHBH). P. mendocina possesses
CC two highly homologous pobA genes (see V84270-71), both of which
CC encode PHBH enzymes (see W88281-82) able to convert

CC para-hydroxybenzoic acid (PHBA) to protocatechuate. A claimed
CC method for the production of PHBA comprises: (a) culturing a
CC poBA(-) Pseudomonas strain in a medium containing an aromatic
CC organic substrate (e.g. toluene or p-cresol), at least one C-source
CC (e.g. glucose or succinate) and an N-source, where the poBA(-)
CC Pseudomonas strain comprises poBA genes encoding the toluene
CC monooxygenase toluene degradation pathway and where the poBA(-)
CC Pseudomonas strain does not produce any detectable PHBH activity
CC owing to poBA gene disruption, so that PHBA accumulates at a rate
CC of about 0.01-1 g PHBA/g.cell.hr; and (b) recovering the PHBA. The
CC PHBA is used as a monomer for liquid crystal polymers. Esters of
CC PHBA can also be used as backbone modifiers in condensation
CC polymers, e.g. polyesters, and are also used to make paraben
CC preservatives.
SQ Sequence 395 AA;

Query Match 48.6%; Score 34; DB 1; Length 395;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy    2 YRLLIKVIR 10
        ||:|:||:|
Db  310 YRILVKVYR 318
```

RESULT 12
P81184
ID P81184 standard; Protein; 1594 AA.
AC P81184;
DT 29-OCT-1992 (first entry)
DE Sequence encoded by the 2nd reading frame of the peplomeric gene
DE of FIVP strain 79-1146.
KW Vaccine; peplomeric protein; E2 gene.
OS Feline infectious peritonitis virus.
PN EP-264979-A.
PD 27-APR-1988.
PF 01-SEP-1987; 201657.
PR 05-SEP-1986; NL-002244.
PA (DUIN) DUPHAR INT RES BV.
PI De Groot RJ, Spaan WJM, Van Der Zeijst BAM;
DR WPI: 88-114147/17.
DR N-PSDB; N81533.
PT Gene for feline infectious peritonitis virus - and gene prod.
PT useful as antigenic protein for vaccine
PS Disclosure; Fig 1; 13pp; English.
CC cDNA was prepd. from FIPv straim 79-1146. N81533 gives the sequence
CC of the peplomeric gene in three reading frames. The top reading
CC frame is an open reading frame of 4356 nucleotides and has a coding
CC capacity for a precursor polypeptide having a mol. wt. of 160,470
CC (1452 AAs). The beginning and the end of the E2 gene are indicated
CC in the FT of N81533. The first 18 N-terminal AAs have a strong
CC hydrophobic character and presumably comprise a cleavable signal
CC peptide. The extreme carboxy-terminal part comprises a region of 20
CC hydrophobic AAs, which presumably serves as a transmembrane anchor.
CC The FIPV peplomeric protein has 35 potential glycosylation sites,
CC of which 22 are in the N-terminal part (pos. 1-790) which corresponds
CC to the S-part of the IBV E2 (see P81183). N.B. IBV - infectious
CC bronchitis virus. "x" in the AA sequence denotes the translation
CC of a stop codon.
SQ Sequence 1594 AA;

Query Match 48.6%; Score 34; DB 1; Length 1594;
Best Local Similarity 57.1%; Pred. No. 3.5e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy    2 YRLLIKVIRIVLKY 15
        |||:| ||||
Db  787 YRLIIAVCTICLCY 800
```

RESULT 13

W76776
ID W76776 standard; Protein; 1745 AA.
AC W76776;
DT 15-JAN-1999 (first entry)
DE D. immitis ankyrin pDIAnk1745 protein.
KW Ankyrin; helminth; parasite; vaccine; therapy; infection;
KW passive immunogen; cytotoxic agent.
OS Dirofilaria immitis.
PN US5827692-A.
PD 27-OCT-1998.
PF 24-APR-1997; 847429.
PR 24-APR-1997; US-847429.
PA (HESK-) HESKA CORP.
PI Blehm ES, Tang L;
WPI: 98-593992/50.
DR N-PSDB; V63020.
PT Nucleic acids encoding ankyrins from helminth parasites - useful for
PT recombinant production of the proteins for use as vaccines and
PT treatments against helminth infection
PS Claim 8; Column 109-118; 84pp; English.
CC W76779-W76777 represent ankyrin proteins isolated from the helminth
CC parasites Dirofilaria immitis and Brugia malayi. The nucleic acids and
CC recombinant products are useful for the recombinant production of the
CC ankyrin polypeptides. These proteins can then be used as vaccines against
CC parasitic helminth, e.g. D. immitis or B. malayi. They can also be used
CC for therapy after infection, and to raise antibodies, also for use in
CC therapeutics, as passive immunogens, or as therapeutics against helminths
CC on conjugation to cytotoxic agents. The nucleic acids contained in
CC viruses, may also be used as viral vaccines, and the nucleic acids
CC themselves or in vectors may be used as genetic vaccines.
SQ Sequence 1745 AA;

Query Match 48.6%; Score 34; DB 1; Length 1745;
Best Local Similarity 50.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 1; Gaps 0;

```
Qy    6 IKVIRIVLKY 15
        ||:: ::|||
Db  393 IKIVELLKY 402
```

RESULT 14
W70608
ID W70608 standard; Protein; 1745 AA.
AC W70608;
DT 21-JAN-1999 (first entry)
DE Full length ankyrin protein.
KW Ankyrin protein; parasitic helminth; heartworm disease;
KW elephantiasis; hydrocele.
OS Dirofilaria immitis.
PN US5824306-A.
PD 20-OCT-1998.
PF 26-FEB-1998; 031485.
PR 24-APR-1997; US-847429.
PR 26-FEB-1998; US-031485.
PA (HESK-) HESKA CORP.
PI Blehm ES, Tang L;
WPI: 98-593373/50.
DR N-PSDB; V63314.
PT Dirofilaria and Brugia ankyrin proteins and antibodies - useful for
PT protection of animals from disease caused by parasitic helminth
PS Claim 5; Columns 107-118; 84pp; English.
CC The present sequence represents a full length Dirofilaria immitis ankyrin
CC protein. The ankyrin protein, or anti-ankyrin antibodies, may be used
CC to protect an animal from disease caused by a parasitic helminth,
CC especially where the disease is heartworm disease, elephantiasis or
CC hydrocele.
SQ Sequence 1745 AA;

Query Match 48.6%; Score 34; DB 1; Length 1745;
Best Local Similarity 50.0%;

```
Matches     5;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6  IKVIRIVLKY 15
           !!::.:!||
Db    393  IKIVELLKY 402

RESULT  15
R92907
ID   R92907 standard; peptide; 20 AA.
AC   R92907;
DT   16-MAY-1996 (first entry)
DE   HLA-B2702 CTL modulating peptide (B2702.84-75/75-84).
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   immunosuppressant; graft versus host disorder; transplantation; therapy;
KW   class I MHC; HLA-B2702.
OS   Synthetic.
PN   WO9526979-A1.
PD   12-OCT-1995.
PF   05-APR-1995; U04349.
PR   05-APR-1994; US-222851.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Clayberger C, Krensky AM, Parham P;
DR   WPI; 95-358582/46.
PT   Extension of acceptance period of transplants from MHC unmatched
PT   donor hosts, - using Class I B75-84 MHC antigen of the recipient
PT   host
PS   Example 15; Page 36; 80pp; English.
CC   R83061-R83085, R83090-R83096 and R92907-R92914 represent fragments of
CC   class I major histocompatibility complex (MHC) antigens.  This sequence
CC   is an inverted dimer of residues 75-84 of the alpha-1 domain of the class
CC   I MHC HLA-B2702.  These sequences can be used to extend the period of
CC   acceptance by a recipient of a transplant from an MHC unmatched donor.
CC   The peptides are administered to a patient in conjunction with a
CC   subtherapeutic amount of an immunosuppressant.  This is administered to
CC   the patient for a limited period of time (compared to the lifetime
CC   administration for current treatments)  The peptides particularly
CC   modulate (or inhibit) the activity of the cytotoxic T lymphocytes (CTLs)
CC   of the patient.
SQ   Sequence   20 AA;

Query Match             47.1%;  Score 33;  DB 1;  Length 20;
Best Local Similarity   40.0%;  Pred. No. 5.5;
Matches    8;  Conservative    4;  Mismatches    4;  Indels    2;  Gaps    1;

Qy      2  YRLLIKV------IRIVLKY 15
           ||| |::    :||| |:|
Db      1  YRLAIRLNERRENLRIALRY 20


Search completed; February  8, 2000, 04:05:40
Job time: 9357 sec
```

GenCore version 4.5
, Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 8, 2000, 00:02:05 ; Search time 75.85 Seconds
(without alignments)
2.541 Million cell updates/sec

Title: US-08-653-294-21
Perfect score: 70
Sequence: 1 AYRLLIKVIRIVLKY 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 130275 seqs, 12848600 residues

Total number of hits satisfying chosen parameters: 130275

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgnl_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgnl_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgnl_6/ptodata/2/iaa/PCTUS9_COMB.pep:*
4: /cgnl_6/ptodata/2/iaa/backfiles1.pep:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	51.4	222	2	US-08-756-771-5	Sequence 5, Appli
2	36	51.4	222	2	US-09-096-571-5	Sequence 5, Appli
3	36	51.4	1226	2	US-08-540-804-12	Sequence 12, Appl
4	36	51.4	1226	2	US-08-218-265-12	Sequence 12, Appl
5	35	50.0	302	2	US-09-031-485-38	Sequence 38, Appl
6	35	50.0	302	2	US-09-031-485-72	Sequence 38, Appl
7	34	48.6	33	2	US-09-031-485-72	Sequence 72, Appl
8	34	48.6	33	2	US-08-847-429A-38	Sequence 72, Appl
9	34	48.6	303	2	US-08-847-429A-72	Sequence 23, Appl
10	34	48.6	303	2	US-09-031-485-23	Sequence 23, Appl
11	34	48.6	1745	2	US-08-847-429A-23	Sequence 33, Appl
12	34	48.6	1745	2	US-09-031-485-33	Sequence 33, Appl
13	33	47.1	20	1	US-08-847-429A-33	Sequence 33, Appl
14	33	47.1	20	3	PCT-US94-12985-1	Sequence 1, Appli
15	33	47.1	20	3	PCT-US94-12985-4	Sequence 4, Appli
16	33	47.1	2860	2	US-08-826-267-2	Sequence 5, Appli
17	32	45.7	12	3	PCT-US94-12985-5	Sequence 5, Appli
18	32	45.7	227	1	US-08-597-236-6	Sequence 6, Appli
19	32	45.7	227	1	US-08-746-682A-6	Sequence 6, Appli
20	32	45.7	384	2	US-08-637-759B-375	Sequence 375, App
21	31	44.3	575	2	US-08-766-858A-5	Sequence 5, Appli
22	31	44.3	101	1	US-07-820-154A-7	Sequence 7, Appli
23	31	44.3	101	1	US-08-220-401-5	Sequence 5, Appli
24	31	44.3	101	1	US-08-437-362-5	Sequence 5, Appli
25	31	44.3	101	2	US-08-097-554A-7	Sequence 7, Appli
26	31	44.3	101	3	PCT-US93-00324-7	Sequence 7, Appli
27	31	44.3	437	2	US-08-538-816A-2	Sequence 2, Appli
28	31	44.3	437	2	US-09-076-651-2	Sequence 2, Appli
29	30	42.9	27	1	US-08-34-427B-1	Sequence 1, Appli
30	30	42.9	27	1	US-08-343-427B-2	Sequence 2, Appli
31	30	42.9	30	1	US-08-343-427B-3	Sequence 3, Appli
32	30	42.9	30	1	US-08-343-427B-4	Sequence 4, Appli
33	30	42.9	30	1	US-08-343-427B-7	Sequence 7, Appli
34	30	42.9	30	1	US-08-343-427B-8	Sequence 8, Appli
35	30	42.9	30	1	US-08-343-427B-9	Sequence 9, Appli
36	30	42.9	30	1	US-08-343-427B-10	Sequence 10, Appl
37	30	42.9	30	2	US-08-652-450A-4	Sequence 4, Appli
38	30	42.9	30	1	US-08-652-450A-2	Sequence 2, Appli
39	30	42.9	35	2	US-08-343-427B-11	Sequence 11, Appl
40	30	42.9	35	2	US-08-652-450A-1	Sequence 1, Appli
41	30	42.9	35	2	US-08-750-194-1	Sequence 1, Appli
42	30	42.9	35	4	5223481-5	Patent No. 5223481
43	30	42.9	35	4	5223481-6	Patent No. 5223481
44	30	42.9	35	4	5223481-7	Patent No. 5223481
45	30	42.9	35	4	5455227-2	Patent No. 5455227

ALIGNMENTS

RESULT 1
US-08-756-771-5
; Sequence 5, Application US/08756771
; Patent No. 5817497
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,771
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 193710
;
US-08-756-771-5

Query Match 51.4%; Score 36; DB 2; Length 222;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 6 IKVIRIVLKY 15

```
Db   213 VEVVRIVLKF 222

RESULT  2
US-09-096-571-5
; Sequence 5, Application US/09096571
; Patent No. 5976528
; GENERAL INFORMATION:
;   APPLICANT: Goll, Surya K.
;   APPLICANT: Hillman, Jennifer L.
;   TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;     STREET: 3174 Porter Drive
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: US
;     ZIP: 94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/096,571
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/756,771
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Billings, Lucy J.
;     REGISTRATION NUMBER: 36,749
;     REFERENCE/DOCKET NUMBER: PF-0162 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-855-0555
;     TELEFAX: 415-845-4166
;     TELEX:
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 222 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: GenBank
;     CLONE: 193710
US-09-096-571-5

Query Match           51.4%;  Score 36;  DB 2;  Length 222;
Best Local Similarity 60.0%;  Pred. No. 16;
Matches    6;  Conservative   4;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     6 IKVIRIVLKY 15
         ::|:||||||
Db   213 VEVVRIVLKF 222

RESULT  3
US-08-540-804-12
; Sequence 12, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
;   APPLICANT: Young, Richard A.
;   APPLICANT: Koleske, Anthony J.
;   APPLICANT: Thompson, Craig M.
;   APPLICANT: Chao, David M.
;   TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
;   TITLE OF INVENTION: Transcription and Methods of Use Therefor
;   NUMBER OF SEQUENCES: 39
```

```
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: Massachusetts
;     COUNTRY: USA
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC Compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/540,804
;     FILING DATE: 11-OCT-1995
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/521,872
;     FILING DATE: 21-AUG-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/218,265
;     FILING DATE: 25-MAR-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: WHI94-03A2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1226 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-540-804-12

Query Match           51.4%;  Score 36;  DB 2;  Length 1226;
Best Local Similarity 77.8%;  Pred. No. 90;
Matches    7;  Conservative   1;  Mismatches   1;  Indels   1;  Gaps   0;

Qy     2 YRLLIKVIR 10
         | |||:|:||
Db   999 YHLLIKIIR 1007

RESULT  4
US-08-218-265-12
; Sequence 12, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
;   APPLICANT: Young, Richard A.
;   APPLICANT: Koleske, Anthony J.
;   APPLICANT: Thompson, Craig M.
;   TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
;   TITLE OF INVENTION: Transcription and Methods of Use Thereof
;   NUMBER OF SEQUENCES: 35
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: MA
;     COUNTRY: US
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC Compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/218,265
;     FILING DATE: 25-MAR-1994
```

```
          CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Granahan, Patricia
;       REGISTRATION NUMBER: 32,227
;       REFERENCE/DOCKET NUMBER: WHI94-03
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 617-861-6240
;       TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 1226 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-218-265-12

Query Match             51.4%;  Score 36;  DB 2;  Length 1226;
Best Local Similarity   77.8%;  Pred. No. 90;
Matches   7;  Conservative   1;  Mismatches   1;  Indels   1;  Gaps   0;

Qy    2  YRLLIKVIR 10
           | |||:||
Db  999  YHLLIKIIR 1007

RESULT 5
US-09-031-485-38
; Sequence 38, Application US/09031485
; Patent No. 5824306
; GENERAL INFORMATION:
;    APPLICANT: Tang, Liang
;    APPLICANT: Blehm, E. Scot
;    TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;    TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION: USES THEREOF
;    NUMBER OF SEQUENCES: 85
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Carol Talkington Verser, Ph.D.
;       ADDRESSEE: Heska Corporation
;       STREET: 1825 Sharp Point Drive
;       CITY: Fort Collins
;       STATE: Colorado
;       COUNTRY: USA
;       ZIP: 80525
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: Windows 95
;       SOFTWARE: WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/031,485
;       FILING DATE:
;       CLASSIFICATION: 530
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US/08/847,429
;       FILING DATE: 24-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Verser, Carol Talkington
;       REGISTRATION NUMBER: 37,459
;       REFERENCE/DOCKET NUMBER: HW-5
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 970/493-7272
;       TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 38:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 302 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-09-031-485-38


Query Match             50.0%;  Score 35;  DB 2;  Length 302;
Best Local Similarity   60.0%;  Pred. No. 33;
Matches   6;  Conservative   3;  Mismatches   3;  Indels   1;  Gaps   0;

Qy   6  IKVIRIVLKY 15
          |||:.:|||
Db  53  IKVVELLKY 62

RESULT 6
US-08-847-429A-38
; Sequence 38, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
;    APPLICANT: Tang, Liang
;    APPLICANT: Blehm, E. Scot
;    TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;    TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION: USES THEREOF
;    NUMBER OF SEQUENCES: 85
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Carol Talkington Verser, Ph.D.
;       ADDRESSEE: Heska Corporation
;       STREET: 1825 Sharp Point Drive
;       CITY: Fort Collins
;       STATE: Colorado
;       COUNTRY: USA
;       ZIP: 80525
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: Windows 95
;       SOFTWARE: WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/847,429A
;       FILING DATE: 24-APR-1997
;       CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Verser, Carol Talkington
;       REGISTRATION NUMBER: 37,459
;       REFERENCE/DOCKET NUMBER: HW-5
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 970/493-7272
;       TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 38:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 302 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-847-429A-38


Query Match             50.0%;  Score 35;  DB 2;  Length 302;
Best Local Similarity   60.0%;  Pred. No. 33;
Matches   6;  Conservative   3;  Mismatches   3;  Indels   1;  Gaps   0;

Qy   6  IKVIRIVLKY 15
          |||:.:|||
Db  53  IKVVELLKY 62

RESULT 7
US-09-031-485-72
; Sequence 72, Application US/09031485
; Patent No. 5824306
; GENERAL INFORMATION:
;    APPLICANT: Tang, Liang
;    APPLICANT: Blehm, E. Scot
;    TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;    TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;    TITLE OF INVENTION: USES THEREOF
;    NUMBER OF SEQUENCES: 85
```

        CORRESPONDENCE ADDRESS:
  ,   ADDRESSE: Carol Talkington Verser, Ph.D.
  ,   ADDRESSE: Heska Corporation
  ,   STREET: 1825 Sharp Point Drive
  ,   CITY: Fort Collins
  ,   STATE: Colorado
  ,   COUNTRY: USA
  ,   ZIP: 80525
        COMPUTER READABLE FORM:
  ,   MEDIUM TYPE: Floppy disk
  ,   COMPUTER: IBM PC compatible
  ,   OPERATING SYSTEM: Windows 95
  ,   SOFTWARE: WordPerfect for Windows, Version 7.0
        CURRENT APPLICATION DATA:
  ,   APPLICATION NUMBER: US/09/031,485
  ,   FILING DATE:
  ,   CLASSIFICATION: 530
        PRIOR APPLICATION DATA:
  ,   APPLICATION NUMBER: US/08/847,429
  ,   FILING DATE: 24-APR-1997
        ATTORNEY/AGENT INFORMATION:
  ,   NAME: Verser, Carol Talkington
  ,   REGISTRATION NUMBER: 37,459
  ,   REFERENCE/DOCKET NUMBER: HW-5
        TELECOMMUNICATION INFORMATION:
  ,   TELEPHONE: 970/493-7272
  ,   TELEFAX: 970/484-9505
  ;   INFORMATION FOR SEQ ID NO: 72:
  ;   SEQUENCE CHARACTERISTICS:
  ,   LENGTH: 33 amino acids
  ,   TYPE: amino acid
  ,   TOPOLOGY: linear
  ;   MOLECULE TYPE: protein
US-09-031-485-72:

Query Match   48.6%; Score 34; DB 2; Length 33;
Best Local Similarity 50.0%; Pred. No. 5.2;
Matches   5; Conservative   4; Mismatches   4; Indels   1; Gaps   0;

Qy   6 IKVIRIVLKY 15
          ||:: ::|||
Db   15 IKIVELLKY 24

RESULT 8
US-08-847-429A-72
; Sequence 72, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
  ;   APPLICANT: Tang, Liang
  ;   APPLICANT: Blehm, E. Scot
  ;   TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
  ;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
  ;   TITLE OF INVENTION: USES THEREOF
  ;   NUMBER OF SEQUENCES: 85
        CORRESPONDENCE ADDRESS:
  ,   ADDRESSE: Carol Talkington Verser, Ph.D.
  ,   ADDRESSE: Heska Corporation
  ,   STREET: 1825 Sharp Point Drive
  ,   CITY: Fort Collins
  ,   STATE: Colorado
  ,   COUNTRY: USA
  ,   ZIP: 80525
        COMPUTER READABLE FORM:
  ,   MEDIUM TYPE: Floppy disk
  ,   COMPUTER: IBM PC compatible
  ,   OPERATING SYSTEM: Windows 95
  ,   SOFTWARE: WordPerfect for Windows, Version 7.0
        CURRENT APPLICATION DATA:
  ,   APPLICATION NUMBER: US/08/847,429A
  ,   FILING DATE: 24-APR-1997
  .   CLASSIFICATION: 435

        ATTORNEY/AGENT INFORMATION:
  ,   NAME: Verser, Carol Talkington
  ,   REGISTRATION NUMBER: 37,459
  ,   REFERENCE/DOCKET NUMBER: HW-5
        TELECOMMUNICATION INFORMATION:
  ,   TELEPHONE: 970/493-7272
  ,   TELEFAX: 970/484-9505
  ;   INFORMATION FOR SEQ ID NO: 72:
  ;   SEQUENCE CHARACTERISTICS:
  ,   LENGTH: 33 amino acids
  ,   TYPE: amino acid
  ,   TOPOLOGY: linear
  ;   MOLECULE TYPE: protein
US-08-847-429A-72:

Query Match   48.6%; Score 34; DB 2; Length 33;
Best Local Similarity 50.0%; Pred. No. 5.2;
Matches   5; Conservative   4; Mismatches   4; Indels   1; Gaps   0;

Qy   6 IKVIRIVLKY 15
          ||:: ::|||
Db   15 IKIVELLKY 24

RESULT 9
US-09-031-485-23
; Sequence 23, Application US/09031485
; Patent No. 584306
; GENERAL INFORMATION:
  ;   APPLICANT: Tang, Liang
  ;   APPLICANT: Blehm, E. Scot
  ;   TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
  ;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
  ;   TITLE OF INVENTION: USES THEREOF
  ;   NUMBER OF SEQUENCES: 85
        CORRESPONDENCE ADDRESS:
  ,   ADDRESSE: Carol Talkington Verser, Ph.D.
  ,   ADDRESSE: Heska Corporation
  ,   STREET: 1825 Sharp Point Drive
  ,   CITY: Fort Collins
  ,   STATE: Colorado
  ,   COUNTRY: USA
  ,   ZIP: 80525
        COMPUTER READABLE FORM:
  ,   MEDIUM TYPE: Floppy disk
  ,   COMPUTER: IBM PC compatible
  ,   OPERATING SYSTEM: Windows 95
  ,   SOFTWARE: WordPerfect for Windows, Version 7.0
        CURRENT APPLICATION DATA:
  ,   APPLICATION NUMBER: US/09/031,485
  ,   FILING DATE:
  ,   CLASSIFICATION: 530
        PRIOR APPLICATION DATA:
  ,   APPLICATION NUMBER: US/08/847,429
  ,   FILING DATE: 24-APR-1997
        ATTORNEY/AGENT INFORMATION:
  ,   NAME: Verser, Carol Talkington
  ,   REGISTRATION NUMBER: 37,459
  ,   REFERENCE/DOCKET NUMBER: HW-5
        TELECOMMUNICATION INFORMATION:
  ,   TELEPHONE: 970/493-7272
  ,   TELEFAX: 970/484-9505
  ;   INFORMATION FOR SEQ ID NO: 23:
  ;   SEQUENCE CHARACTERISTICS:
  ,   LENGTH: 303 amino acids
  ,   TYPE: amino acid
  ,   TOPOLOGY: linear
  ;   MOLECULE TYPE: protein
US-09-031-485-23:

Query Match   48.6%; Score 34; DB 2; Length 303;

```
Best Local Similarity   50.0%;   Pred. No. 49;
Matches    5; Conservative    4; Mismatches    1; Indels    0; Gaps    0;

Qy    6 IKVIRIVLKY 15
        ||::.::|||
Db   54 IKIVELLLKY 63

RESULT 10
US-08-847-429A-23
; Sequence 23, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
;   APPLICANT: Tang, Liang
;   APPLICANT: Blehm, E. Scot
;   TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;   TITLE OF INVENTION: USES THEREOF
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Carol Talkington Verser, Ph.D.
;     ADDRESSEE: Heska Corporation
;     STREET: 1825 Sharp Point Drive
;     CITY: Fort Collins
;     STATE: Colorado
;     COUNTRY: USA
;     ZIP: 80525
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: Windows 95
;     SOFTWARE: WordPerfect for Windows, Version 7.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/847,429A
;     FILING DATE: 24-APR-1997
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Verser, Carol Talkington
;     REGISTRATION NUMBER: 37,459
;     REFERENCE/DOCKET NUMBER: HW-5
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 970/493-7272
;     TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO:  23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  303 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-847-429A-23


Query Match         48.6%;    Score 34;  DB 2;  Length 303;
Best Local Similarity   50.0%;   Pred. No. 49;
Matches    5; Conservative    4; Mismatches    1; Indels    0; Gaps    0;

Qy    6 IKVIRIVLKY 15
        ||::.::|||
Db   54 IKIVELLLKY 63

RESULT 11
US-09-031-485-33
; Sequence 33, Application US/09031485
; Patent No. 5824306
; GENERAL INFORMATION:
;   APPLICANT: Tang, Liang
;   APPLICANT: Blehm, E. Scot
;   TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;   TITLE OF INVENTION: USES THEREOF
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:


;     ADDRESSEE: Carol Talkington Verser, Ph.D.
;     ADDRESSEE: Heska Corporation
;     STREET: 1825 Sharp Point Drive
;     CITY: Fort Collins
;     STATE: Colorado
;     COUNTRY: USA
;     ZIP: 80525
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: Windows 95
;     SOFTWARE: WordPerfect for Windows, Version 7.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/031,485
;     FILING DATE:
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/847,429
;     FILING DATE: 24-APR-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Verser, Carol Talkington
;     REGISTRATION NUMBER: 37,459
;     REFERENCE/DOCKET NUMBER: HW-5
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 970/493-7272
;     TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO:  33:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1745 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-09-031-485-33


Query Match         48.6%;    Score 34;  DB 2;  Length 1745;
Best Local Similarity   50.0%;   Pred. No. 2.9e+02;
Matches    5; Conservative    4; Mismatches    1; Indels    0; Gaps    0;

Qy    6 IKVIRIVLKY 15
        ||::.::|||
Db  393 IKIVELLLKY 402

RESULT 12
US-08-847-429A-33
; Sequence 33, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
;   APPLICANT: Tang, Liang
;   APPLICANT: Blehm, E. Scot
;   TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
;   TITLE OF INVENTION: USES THEREOF
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Carol Talkington Verser, Ph.D.
;     ADDRESSEE: Heska Corporation
;     STREET: 1825 Sharp Point Drive
;     CITY: Fort Collins
;     STATE: Colorado
;     COUNTRY: USA
;     ZIP: 80525
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: Windows 95
;     SOFTWARE: WordPerfect for Windows, Version 7.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/847,429A
;     FILING DATE: 24-APR-1997
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
```

```
                  NAME: Verser, Carol Talkington
                  REGISTRATION NUMBER: 37,459
                  REFERENCE/DOCKET NUMBER: HW-5
                  TELECOMMUNICATION INFORMATION:
                  TELEPHONE: 970/493-7272
                  TELEFAX: 970/484-9505
       INFORMATION FOR SEQ ID NO: 33:
              SEQUENCE CHARACTERISTICS:
                  LENGTH: 1745 amino acids
                  TYPE: amino acid
                  TOPOLOGY: linear
                  MOLECULE TYPE: protein
US-08-847-429A-33

Query Match          48.6%;  Score  34;  DB 2;  Length 1745;
Best Local Similarity 50.0%;  Pred. No. 2.9e+02;
Matches   5;  Conservative   4;  Mismatches   4;  Indels   1;  Gaps    0;

Qy    6  IKVIRIVLKY  15
         ||::.:|||
Db  393  IKIVELLKY  402

RESULT 13
US-08-222-851-33
; Sequence 33, Application US/08222851
; Patent No. 5723128
  GENERAL INFORMATION:
       APPLICANT: CLAYBERGER, CAROL A.
       APPLICANT: KRENSKY, ALAN M.
       APPLICANT: PARHAM, PETER
       TITLE OF INVENTION: CYTOTOXIC T-CELL LYMPHOCYTE ("CTL")
       TITLE OF INVENTION: ACTIVITY REGULATION BY CLASS I MHC PEPTIDES
       NUMBER OF SEQUENCES: 43
       CORRESPONDENCE ADDRESS:
              ADDRESSEE: MORRISON & FOERSTER
              STREET: 2000 PENNSYLVANIA AVENUE, NW, STE 5500
              CITY: WASHINGTON
              STATE: DC
              COUNTRY: USA
              ZIP: 20006-1812
       COMPUTER READABLE FORM:
              MEDIUM TYPE: Floppy disk
              COMPUTER: IBM PC compatible
              OPERATING SYSTEM: PC-DOS/MS-DOS
              SOFTWARE: PatentIn Release #1.0, Version #1.30
       CURRENT APPLICATION DATA:
              APPLICATION NUMBER: US/08/222,851
              FILING DATE: 05-APR-1994
              CLASSIFICATION: 424
       ATTORNEY/AGENT INFORMATION:
              NAME: MILLMAN, ROBERT A.
              REGISTRATION NUMBER: 36,217
       TELECOMMUNICATION INFORMATION:
              REFERENCE/DOCKET NUMBER: 28600-20200.22
              TELEPHONE: (202) 887-1500
              TELEFAX: (202) 494-0792
              TELEX: 90-4030 MRSNFOERSWSH
       INFORMATION FOR SEQ ID NO: 33:
              SEQUENCE CHARACTERISTICS:
              LENGTH: 20 amino acids
              TYPE: amino acid
              STRANDEDNESS: single
              TOPOLOGY: linear
US-08-222-851-33

Query Match          47.1%;  Score  33;  DB 1;  Length 20;
Best Local Similarity 40.0%;  Pred. No. 4.7;
Matches   8;  Conservative   4;  Mismatches   4;  Indels   2;  Gaps    6;

Qy    2  YRLLIKV------IRIVLKY  15
```

```
         ||| |::   .:|| |:|
Db    1  YRLAIRLNERRENLRIALRY  20

RESULT 14
PCT-US94-12985-1
; Sequence 1, Application PCT/US9412985
  GENERAL INFORMATION:
       APPLICANT: The Board of Trustees for the Leland Stanford Junior
       APPLICANT: University
       TITLE OF INVENTION: SURFACE MEMBRANE PROTEINS AND THEIR
       TITLE OF INVENTION: EFFECT ON IMMUNE RESPONSE
       NUMBER OF SEQUENCES: 20
       CORRESPONDENCE ADDRESS:
              ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
              STREET: 4 Embarcadero Center, Suite 3400
              CITY: San Francisco
              STATE: California
              COUNTRY: USA
              ZIP: 9411-4187
       COMPUTER READABLE FORM:
              MEDIUM TYPE: Floppy disk
              COMPUTER: IBM PC compatible
              OPERATING SYSTEM: PC-DOS/MS-DOS
              SOFTWARE: PatentIn Release #1.0, Version #1.25
       CURRENT APPLICATION DATA:
              APPLICATION NUMBER: PCT/US94/12985
              FILING DATE: 10-NOV-1994
              CLASSIFICATION:
       PRIOR APPLICATION DATA:
              APPLICATION NUMBER: 08/150,493
              FILING DATE: 10-NOV-1993
       ATTORNEY/AGENT INFORMATION:
              NAME: Rowland, Bertram I
              REGISTRATION NUMBER: 20,015
              REFERENCE/DOCKET NUMBER: FP-58976-PC/BIR
       TELECOMMUNICATION INFORMATION:
              TELEPHONE: (415) 781-1989
              TELEFAX: (415) 398-3249
              TELEX: 910 277299
       INFORMATION FOR SEQ ID NO: 1:
              SEQUENCE CHARACTERISTICS:
              LENGTH: 20 amino acids
              TYPE: amino acid
              STRANDEDNESS: single
              TOPOLOGY: linear
              MOLECULE TYPE: peptide
PCT-US94-12985-1

Query Match          47.1%;  Score  33;  DB 3;  Length 20;
Best Local Similarity 40.0%;  Pred. No. 4.7;
Matches   8;  Conservative   4;  Mismatches   4;  Indels   2;  Gaps    1;

Qy    2  YRLLIKV-----IRIVLKY  15
         ||| |::| |:|
Db    1  YRLAIRLNERRENLRIALRY  20

RESULT 15
PCT-US94-12985-4
; Sequence 4, Application PCT/US9412985
  GENERAL INFORMATION:
       APPLICANT: The Board of Trustees for the Leland Stanford Junior
       APPLICANT: University
       TITLE OF INVENTION: SURFACE MEMBRANE PROTEINS AND THEIR
       TITLE OF INVENTION: EFFECT ON IMMUNE RESPONSE
       NUMBER OF SEQUENCES: 20
       CORRESPONDENCE ADDRESS:
              ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
              STREET: 4 Embarcadero Center, Suite 3400
              CITY: San Francisco
              STATE: California
```

```
COUNTRY: USA
  ZIP: 94111-4187
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: PCT/US94/12985
  FILING DATE: 10-NOV-1994
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/150,493
  FILING DATE: 10-NOV-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Rowland, Bertram I
  REGISTRATION NUMBER: 20,015
  REFERENCE/DOCKET NUMBER: FP-58976-PC/BIR
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 781-1989
  TELEFAX: (415) 398-3249
  TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
  LENGTH: 20 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: peptide
PCT-US94-12985-4
```

```
Query Match        47.1%;    Score 33;  DB 3;  Length 20;
Best Local Similarity   40.0%;   Pred. No. 4.7;
Matches    8;  Conservative    4;  Mismatches    2;  Indels    6;  Gaps    1;

Qy     2 YRLLIKV-----IRIVLKY 15
         ||| ||::     ||| ||:|
Db     1 YRLAIRLNERRENLRIALRY 20
```

```
Search completed: February  8, 2000, 01:01:46
Job time: 3581 sec
```

THIS PAGE BLANK (USPTO)

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February 7, 2000, 11:54:34 ; Search time 111.22 seconds
                                           (without alignments)
                                           6.362 Million cell updates/sec

Title:       US-08-653-294-21
Perfect score: 70
Sequence:    1 AYRLLIKVIRIVLKY 15

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    142080 seqs, 47169319 residues

Total number of hits satisfying chosen parameters:        142080

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
             Listing first 45 summaries

Database :   PIR_62:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	39.5	56.4	746	2	S67203	probable membrane
2	38	54.3	404	2	F64238	hypothetical prote
3	38	54.3	709	1	OXCKPM	acyl-CoA oxidase (
4	38	54.3	771	1	T16254	hypothetical prote
5	37	52.9	98	2	C32227	hypothetical prote
6	37	52.9	242	2	G75029	hypothetical prote
7	37	52.9	379	2	G75170	hypothetical prote
8	36.5	52.1	724	2	S57604	probable membrane
9	36	51.4	131	2	G70326	hypothetical prote
10	36	51.4	186	2	C71601	probable integral
11	36	51.4	187	2	E71464	probable rRNA meth
12	36	51.4	222	2	S27234	glutathione transf
13	36	51.4	282	2	B69260	hypothetical prote
14	36	51.4	492	2	S65238	probable membrane
15	36	51.4	705	2	S38066	probable finger pr
16	36	51.4	1427	2	S74293	SRB8 protein - yea
17	36	51.4	2048	2	2LNZSE	genome polyprotein
18	36	51.4	2228	1	2LNZSV	genome polyprotein
19	35	50.0	341	2	A72674	hypothetical prote
20	35	50.0	485	2	T09974	H+-transporting AT
21	35	50.0	486	2	B70775	probable atpD prot
22	35	50.0	502	1	OXCKAX	acyl-CoA oxidase (
23	35	50.0	709	1	OXCKX4	acyl-CoA oxidase (
24	35	50.0	709	1	OXCKX	acyl-CoA oxidase (
25	35	50.0	791	2	T02583	hypothetical prote
26	35	50.0	1784	2	A49420	tuberous sclerosis
27	35	50.0	1786	2	A57282	ankyrin-related pr
28	35	50.0	1809	2	T15345	ankyrin-related un
29	35	50.0	1809	2	S57329	tuberous sclerosis
30	35	50.0	1815	2	T15346	elegans ankyrin-re
31	35	50.0	1867	2	T15344	ankyrin-related un
32	35	50.0	2039	2	T15347	ankyrin-related un
33	35	50.0	2376	2	S48405	probable membrane
34	34	48.6	3973	2	B71612	hypothetical prote
35	34	48.6	180	2	G72616	hypothetical prote
36	34	48.6	185	2	D72118	hypothetical prote
37	34	48.6	277	2	JC2565	replication protei
38	34	48.6	326	2	S59101	NADH dehydrogenase
39	34	48.6	359	2	H71343	hypothetical prote
40	34	48.6	369	2	T15213	hypothetical prote
41	34	48.6	376	2	H71689	probable UDP-n-ace
42	34	48.6	395	2	S74051	hypothetical prote
43	34	48.6	448	2	S56260	probable membrane
44	34	48.6	651	2	G71697	probable soluble l
45	33	47.1	111	2	A72079	hypothetical prote

ALIGNMENTS

```
RESULT    1
S67203
probable membrane protein YOR299w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O5635
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 06-Feb-1998
C:Accession: S67203
R:Cziepluch, C.; Jauniaux, J.C.; Kordes, E.; Poirey, R.; Pujol, A.; Tobiasch, E.
submitted to the Protein Sequence Database; July 1996
A:Reference number: S67194
A:Accession: S67203
A:Molecule type: DNA
A:Residues: 1-746 <CZI>
A:Cross-references: EMBL:Z75207; NID:g1420661; PID:e252136; PID:g1420662; MIPS:YOR299
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:BUD7
A:Cross-references: SGD:S0005825; MIPS:YOR299w
A:Map position: 15R
C:Keywords: transmembrane protein
F;328-344/Domain: transmembrane #status predicted <TMM>

Query Match             56.4%;  Score 39.5;  DB 2;  Length 746;
Best Local Similarity   45.0%;  Pred. No. 17;
Matches    9; Conservative    5;  Mismatches    5;  Indels    1;  Gaps    1;

Qy      1 AYRLLIKVIRI----VLKY 15
            ||||| :::::|     :|||
Db    502 AYRLLTFIVQITFGWENLLKY 521


RESULT    2
F64238
hypothetical protein MG349 - Mycoplasma genitalium (SGC3)
C;Species: Mycoplasma genitalium
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 10-Oct-1997
C:Accession: F64238
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
  C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: F64238
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-404 <TIGR>
A:Cross-references: GB:U39719; GB:L43967; NID:g1046055; PID:g1046056; TIGR:MG349
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
```

```
Query Match            54.3%;  Score 38;  DB 2;  Length 404;
Best Local Similarity  57.1%;  Pred. No. 18;
Matches  8;  Conservative  3;  Mismatches  3;  Indels  0;  Gaps  0;

Qy    2 YRLLIKVIRIVLKY 15
        |:||| ||| :||
Db  322 YQLLDKVINLVLDF 335

RESULT 3
OXCKPM
acyl-CoA oxidase (EC 1.3.3.6) PXP4, peroxisomal - yeast (Candida maltosa)
C;Species: Candida maltosa
C;Date:.31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 17-Feb-1995
C;Accession: A29441
R;Hill, D.E.; Boulay, R.; Rogers, D.
Nucleic Acids Res. 16, 365-356, 1988
A;Title: Complete nucleotide sequence of the peroxisomal acyl CoA oxidase from the alkan
A;Reference number: A29441; MUID:88124223
A;Accession: A29441
A;Molecule type: DNA
A;Residues: 1-709 <HIL>
A;Experimental source: ATCC 20184
C;Genetics:
A;Gene: POX4
C;Superfamily: acyl-CoA oxidase
C;Keywords: FAD; fatty acid oxidation; flavoprotein; oxidoreductase; peroxisome

Query Match            54.3%;  Score 38;  DB 1;  Length 709;
Best Local Similarity  46.7%;  Pred. No. 31;
Matches  7;  Conservative  4;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    1 AYRLLIKVIRIVLKY 15
        :||:| :| |:|
Db  307 SYRMLARVSTIALRY 321

RESULT 4
T16254
hypothetical protein F35C8.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16254
R;Wu, X.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F35C8.
A;Reference number: Z18486
A;Accession: T16254
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-771 <WUX>
A;Cross-references: EMBL:U40941; NID:g1072184; PID:g1072187; PIDN:AAA81709.1; CESP:F35C8
A;Gene: CESP:F35C8.7
A;Introns: 87/2; 143/2; 238/1; 289/2; 428/2; 629/3; 678/1; 712/1; 736/1

Query Match            54.3%;  Score 38;  DB 2;  Length 771;
Best Local Similarity  45.5%;  Pred. No. 33;
Matches  5;  Conservative  7;  Mismatches  6;  Indels  0;  Gaps  0;

Qy    5 LIKVIRIVLKY 15
        :||:||:|
Db  578 IVKIIRVLLEY 588

RESULT 5
C32227
hypothetical protein 3 (mer operon) - Bacillus sp.
C;Species: Bacillus sp.
C;Date: 08-Sep-1989 #sequence_revision 08-Sep-1989 #text_change 09-Sep-1997
```

```
C;Accession: C32227
R;Wang, Y.; Moore, M.; Levinson, H.S.; Silver, S.; Walsh, C.; Mahler, I.
J. Bacteriol. 171, 83-92, 1989
A;Title: Nucleotide sequence of a chromosomal mercury resistance determinant from a B
A;Reference number: A32227; MUID:89123092
A;Accession: C32227
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-98 <WAN>
A;Cross-references: GB:M22708; NID:g143188; PID:g1129095

Query Match            52.9%;  Score 37;  DB 2;  Length 98;
Best Local Similarity  58.3%;  Pred. No. 7.6;
Matches  7;  Conservative  2;  Mismatches  3;  Indels  0;  Gaps  0;

Qy    2 YRLLIKVIRIVL 13
        |||| :: |||
Db   46 YRLLFSIVTIVL 57

RESULT 6
G75029
hypothetical protein PAB1160 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: G75029
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: G75029
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <KAW>
A;Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50677.1; PID:e151
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1160

Query Match            52.9%;  Score 37;  DB 2;  Length 242;
Best Local Similarity  60.0%;  Pred. No. 17;
Matches  9;  Conservative  1;  Mismatches  5;  Indels  0;  Gaps  0;

Qy    1 AYRLLIKVIRIVLKY 15
        |||:|| :|| |
Db    8 AYRLIFVIFAFLSY 22

RESULT 7
G75170
hypothetical protein PAB2010 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: G75170
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: G75170
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-379 <KAW>
A;Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49446.1; PID:e151
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB2010

Query Match            52.9%;  Score 37;  DB 2;  Length 379;
Best Local Similarity  77.8%;  Pred. No. 26;
Matches  7;  Conservative  2;  Mismatches  0;  Indels  0;  Gaps  0;
```

Qy   1 AYRLLIKVI 9
         ||:||:|||
Db   19 AYKLLVKVI 27

RESULT 8
S57604
probable membrane protein YMR237w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YM9959.19
C:Species: Saccharomyces cerevisiae
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Nov-1997
C:Genetics:
A:Accession: S57604
R:Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57587
A:Accession: S57604
A:Molecule type: DNA
A:Residues: 1-724 <SKE>
A:Cross-references: EMBL:Z49939; NID:g887599; PID:g887618; MIPS:YMR237w
A:Experimental source: strain AB972
C:Genetics:
A:Map position: 13R
C:Keywords: transmembrane protein
F:666-682/Domain: transmembrane protein #status predicted <TMM>

Query Match   52.1%;   Score 36.5; DB 2; Length 724;
Best Local Similarity   40.0%;   Pred. No. 59;
Matches   8; Conservative   6; Mismatches   1; Indels   5; Gaps   1;

Qy   1 AYRLLIKVIRI----VLKY 15
         ||:|| :::|:| :|||
Db   480 AYKLLTEIVQITGWEQLLKY 499

RESULT 9
G70326
hypothetical protein aq_293 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C:Accession: G70326
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
v.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: G70326
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-131 <AQF>
A:Cross-references: GB:AE000682; NID:g2982979; PID:g2982988; GB:AE000657
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_293
C:Superfamily: Aquifex aeolicus hypothetical protein aq_293

Query Match   51.4%;   Score 36; DB 2; Length 131;
Best Local Similarity   28.6%;   Pred. No. 15;
Matches   4; Conservative   8; Mismatches   2; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLKY 15
         :::|:||:|:::|
Db   46 HKILVKVLNVIIIY 59

RESULT 10
C71601
probable integral membrane protein PFB0985c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: C71601

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: C71601
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-186 <GAR>
A:Cross-references: GB:AE001431; GB:AE001352; NID:g3845327; PID:g3845329; TIGR:PFB098
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0985c

Query Match   51.4%;   Score 36; DB 2; Length 186;
Best Local Similarity   53.3%;   Pred. No. 21;
Matches   8; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
         |:|||:||:||:|:|
Db   145 AFILLIFIIHIVARY 159

RESULT 11
E71464
probable rRNA methylase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
C:Accession: E71464
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: E71464
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <ARN>
A:Cross-references: GB:AE001356; GB:AE001273; NID:g3329299; PID:g3329300
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: yggH

Query Match   51.4%;   Score 36; DB 2; Length 187;
Best Local Similarity   42.9%;   Pred. No. 21;
Matches   6; Conservative   3; Mismatches   5; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLKY 15
         |:|| | | :||:|||
Db   130 YNYLVNAITVMLKY 143

RESULT 12
S27234
glutathione transferase (EC 2.5.1.18) 5.7 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Jun-1999
C:Accession: S27234; S27234; S17542; S17681
R:Zimniak, P.; Eckles, M.A.; Saxena, M.; Awasthi, Y.C.
FEBS Lett. 313, 173-176, 1992
A:Title: A subgroup of class alpha glutathione S-transferases. Cloning of cDNA for mo
A:Reference number: S27234; MUID:93050245
A:Accession: S27234
A:Molecule type: mRNA
A:Residues: 1-222 <ZIM>
A:Cross-references: GB:L06047; NID:g193709; PIDN:AAA37754.1; PID:g193710
R:Medh, R.D.; Saxena, M.; Singhal, S.S.; Ahmad, H.; Awasthi, Y.C.
Biochem. J. 278, 793-799, 1991
A:Title: Characterization of a novel glutathione S-transferase isoenzyme from mouse l
A:Reference number: S17542; MUID:91378941
A:Accession: S17542
A:Molecule type: protein

A;Residues: 106-114,'P',116-120 <MED>
A;Experimental source: lung
A;Accession: S17681
A;Molecule type: protein
A;Residues: 107-113;'G',168-178,'GE',181-184,'X',186-186 <MEW>
A;Experimental source: liver
C;Superfamily: glutathione transferase
C;Keywords: transferase

Query Match　　　　51.48%;　Score 36;　DB 2;　Length 222;
Best Local Similarity　60.0%;　Pred. No. 25;
Matches　6;　Conservative　4;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　6 IKVIRIVLKY 15
　　　　:|:|||||:
Db　213 VEVRRIVLKF 222

RESULT 13
B69260
hypothetical protein AF0082 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C;Accession: B69260
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69260; MUID:98049343
A;Accession: B69260
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-282 <KLE>
A;Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PID:g2650567; TIGR:AF0082

Query Match　　　　51.48%;　Score 36;　DB 2;　Length 282;
Best Local Similarity　53.8%;　Pred. No. 31;
Matches　7;　Conservative　4;　Mismatches　0;　Indels　2;　Gaps　0;

Qy　　2 YRLLIKVIRIVLK 14
　　　　|| |:||:|| ||
Db　269 YRHLMRLIRMTLK 281

RESULT 14
S65238
probable membrane protein YPL219w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein P1745
C;Species: Saccharomyces cerevisiae
C;Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 06-Feb-1998
C;Accession: S65238
R;Rieger, M.; Mueller-Auer, S.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65202
A;Accession: S65238
A;Molecule type: DNA
A;Residues: 1-492 <RIE>
A;Cross-references: EMBL:Z73575; NID:g1370453; PID:g1370454; MIPS:YPL219w
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: SGD:PCL8
C;Cross-references: SGD:S0006140; MIPS:YPL219w
A;Map position: 16L
C;Keywords: transmembrane protein
F:380-396/Domain: transmembrane #status predicted <TMM>

Query Match　　　　51.48%;　Score 36;　DB 2;　Length 492;

Best Local Similarity　50.0%;　Pred. No. 51;
Matches　7;　Conservative　3;　Mismatches　4;　Indels　0;　Gaps　0;

Qy　　1 AYRLLIKVIRIVLK 14
　　　　 |:|:||　||| |
Db　416 AHRIISTIRIATK 429

RESULT 15
S38066
probable finger protein YKL222c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein F705
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 24-Sep-1999
C;Accession: S38066; S44510; S43548
R;Alexandraki, D.; Horaitis, O.; Tzermia, M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38065
A;Accession: S38066
A;Molecule type: DNA
A;Residues: 1-705 <ALE>
A;Cross-references: EMBL:Z28222; NID:g486398; PIDN:CAA82067.1.; PID:g486399; MIPS:YKL2
A;Experimental source: strain S288C
R;Alexandraki, D.; Tzermia, M.
Yeast 10(Suppl.A), S81-S91, 1994
A;Title: Sequencing of a 13.2 kb segment next to the left telomere of yeast chromosom
V.
A;Reference number: S44508; MUID:94378726
A;Accession: S44510
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-705 <AL2>
A;Cross-references: EMBL:X75950; NID:g473124; PIDN:CAA53551.1; PID:g473127
A;Experimental source: strain S288C
C;Genetics:
A;Map position: 11L
C;Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluste
C;Keywords: DNA binding; nucleus; zinc finger
F:19-57/Domain: GAL4 zinc binuclear cluster homology <GAL4>
F:24-52/Region: zinc finger

Query Match　　　　51.48%;　Score 36;　DB 2;　Length 705;
Best Local Similarity　46.7%;　Pred. No. 71;
Matches　7;　Conservative　4;　Mismatches　4;　Indels　0;　Gaps　0;

Qy　　1 AYRLLIKVIRIVLKY 15
　　　　 |:|:||:: |:||!
Db　541 AFRALIQIYTIFLQY 555

Search completed: February 7, 2000, 18:04:35
Job time: 22201 sec

```
              GenCore version 4.5
     Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   February 8, 2000, 01:00:00 ; Search time 68.63 Seconds
                                        (without alignments)
                                        6.527 Million cell updates/sec

Title:          US-08-653-294-21
Perfect score:  70
Sequence:       1 AYRLLIKVIRIVLKY 15

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       82229 seqs, 29864866 residues

Total number of hits satisfying chosen parameters:   82229

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :      SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	% Match	Query Length	DB	ID	Description
1	38	54.3	404	1	Y349_MYCGE	P47591 mycoplasma
2	38	54.3	708	1	CAO4_CANMA	P05335 candida mal
3	37	52.9	390	1	Y4RI_RHISN	P55642 rhizobium s
4	36.5	52.1	724	1	YM76_YEAST	Q05029 saccharomyc
5	36	51.4	222	1	GTA4_MOUSE	P24472 mus musculu
6	36	51.4	705	1	YKW2_YEAST	P35995 saccharomyc
7	36	51.4	1427	1	SRB8_YEAST	P25648 saccharomyc
8	36	51.4	2048	1	RRPL_SENDB	P06829 sendai viru
9	36	51.4	2228	1	RRPL_SEND5	P27566 sendai viru
10	36	51.4	2228	1	RRPL_SENDF	Q06996 sendai viru
11	36	51.4	2228	1	RRPL_SENDZ	P06447 sendai viru
12	35	50.0	485	1	ATPB_MYCLE	P45823 mycobacteri
13	35	50.0	486	1	ATPB_MYCTU	Q10593 mycobacteri
14	35	50.0	502	1	CAO3_CANTR	P11355 candida tro
15	35	50.0	708	1	CAO2_CANTR	P06598 candida tro
16	35	50.0	1807	1	TSC2_HUMAN	P49815 homo sapien
17	35	50.0	1809	1	TSC2_RAT	P49816 rattus norv
18	35	50.0	2376	1	YIM9_YEAST	P40468 saccharomyc
19	34	48.6	227	1	UREF_ACTPL	O54423 actinobacil
20	34	48.6	326	1	NU1M_CHOCR	P48898 chondrus cr
21	34	48.6	448	1	YFH5_YEAST	P43589 saccharomyc
22	33	47.1	221	1	YM17_PARTE	P15618 paramecium
23	33	47.1	254	1	YC8A_METJA	P81318 methanococc
24	33	47.1	517	1	MBI3_YEAST	Q9ZW7 saccharomyc
25	33	47.1	661	1	CAO1_CANTR	P08790 candida tro
26	33	47.1	677	1	RN14_YEAST	P25298 saccharomyc
27	33	47.1	1370	1	VCAP_HCMVA	P16729 human cytom
28	33	47.1	1391	1	LYS2_CANAL	Q12572 candida alb
29	33	47.1	1591	1	TIAM_MOUSE	Q60610 mus musculu
30	33	47.1	1867	1	MOT1_YEAST	P32333 saccharomyc
31	33	47.1	3655	1	YAMB_SCHPO	Q10064 schizosacch
32	32	45.7	130	1	RS8_PYRHO	O59432 pyrococcus
33	32	45.7	133	1	RS8_SULAC	O05636 sulfolobus
34	32	45.7	174	1	YZ19_METJA	Q60279 methanococc
35	32	45.7	184	1	RL5_THEMA	P38517 thermotoga
36	32	45.7	201	1	YIGP_ECOLI	P27852 escherichia
37	32	45.7	283	1	CG1C_CHICK	P55168 gallus gall
38	32	45.7	283	1	CG1C_MOUSE	Q62447 mus musculu
39	32	45.7	298	1	CG1C_RAT	P39947 rattus norv
40	32	45.7	303	1	CG1C_HUMAN	P24863 homo sapien
41	32	45.7	314	1	SSAK_SALTY	P74853 salmonella
42	32	45.7	323	1	ACCA_AQUAE	O67260 aquifex aeo
43	32	45.7	362	1	VP61_MRDV	P22119 maize rough
44	32	45.7	399	1	LICH_HUMAN	P38571 homo sapien
45	32	45.7	400	1	AFC3_ARATH	P51568 arabidopsis

ALIGNMENTS

```
RESULT 1
Y349_MYCGE
ID   Y349_MYCGE     STANDARD;      PRT;   404 AA.
AC   P47591;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   HYPOTHETICAL PROTEIN MG349.
GN   MG349.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE; 96026346.
RA   FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA   FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA   FRITCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA   NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA   TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA   PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U39716; AAC71574.1; -.
DR   TIGR; MG349; -.
KW   Hypothetical protein.
SQ   SEQUENCE   404 AA;  47774 MW;  BEEC5F2B CRC32;

Query Match             54.3%;  Score 38;  DB 1;  Length 404;
Best Local Similarity   57.1%;  Pred. No. 9.8;
Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 YRLLIKVIRIVLKY 15
                |:|| ||| :|| :
Db        322 YQLLDKVINLVLDF 335

RESULT 2
CAO4_CANMA
ID   CAO4_CANMA     STANDARD;      PRT;   708 AA.
AC   P05335;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-MAY-1992 (Rel. 22, Last annotation update)
DE   ACYL-COENZYME A OXIDASE POX4 (EC 1.3.6) (ACYL-COA OXIDASE) (AOX).
GN   POX4.
```

```
OS   Candida maltosa (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Candidaceae; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20184;
RX   MEDLINE; 88124223.
RA   HILL D.E., BOULAY R., ROGERS D.;
RT   "Complete nucleotide sequence of the peroxisomal acyl CoA oxidase
RT   from the alkane-utilizing yeast Candida maltosa.";
RL   Nucleic Acids Res. 16:365-366(1988).
CC   --!-- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC         H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC         FROM 8 TO 18).
CC   --!-- COFACTOR: FAD.
CC   --!-- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC         SYSTEM.
CC   --!-- SUBUNIT: HOMOOCTAMER.
CC   --!-- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X06721; CAA29901.1; -.
DR   PIR; A29441; OXCKPM.
KW   Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW   Multigene family.
FT   INIT_MET       0      0       BY SIMILARITY.
SQ   SEQUENCE   708 AA;  78242 MW;  D5E344D2 CRC32;

Query Match             54.3%;  Score 38;  DB 1;  Length 708;
Best Local Similarity   46.7%;          Pred. No. 17;
Matches    7;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AYRLLIKVIRIVLKY 15
           :||:| :| :| |:|
Db     306 SYRMLARVSTIALRY 320


RESULT   3
Y4RI_RHISN
ID   Y4RI_RHISN     STANDARD;      PRT;     390 AA.
AC   P55642;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 44.0 KD PROTEIN Y4RI.
GN   Y4RI.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97305956.
RA   FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA   PERRET X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   --!-- SIMILARITY: NONE OBVIOUS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000094; AAB91834.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   390 AA;  43978 MW;  F0AF4E11 CRC32;

Query Match             52.9%;  Score 37;  DB 1;  Length 390;
Best Local Similarity   50.0%;          Pred. No. 14;
Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 YRLLIKVIRIVLKY 15
           || |:|:|: | ||
Db     157 YRELLKIARTTLSY 170


RESULT   4
YM76_YEAST
ID   YM76_YEAST     STANDARD;      PRT;     724 AA.
AC   Q05029;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 82.0 KD PROTEIN IN RNA1-RNT1 INTERGENIC REGION.
GN   YMR237W OR YM9959.19.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   SKELTON J., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   --!-- SIMILARITY: TO YEAST CSD3/CHS6 AND YEAST YKR027W.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z49939; CAA90208.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   724 AA;  82048 MW;  6CE22AD2 CRC32;

Query Match             52.1%;  Score 36.5;  DB 1;  Length 724;
Best Local Similarity   40.0%;          Pred. No. 33;
Matches    8;  Conservative    6;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       1 AYRLLIKVIRI----VLKY 15
           |:|| ||     :|||
Db     480 AYKLLFIVQITGWEQLLKY 499


RESULT   5
GTA4_MOUSE
ID   GTA4_MOUSE     STANDARD;      PRT;     222 AA.
AC   P24472;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   GLUTATHIONE S-TRANSFERASE 5.7 (EC 2.5.1.18) (GST 5.7) (GST CLASS-
DE   ALPHA) (GST A4-4) (GSTA4-4).
GN   GSTA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
```

RX   MEDLINE; 93050245.
RA   ZIMNIAK P., ECKLES M.A., SAXENA M., AWASTHI Y.C.;
RT   "A subgroup of class alpha glutathione S-transferases. Cloning of
RT   cDNA for mouse lung glutathione S-transferase GST 5.7.";
RL   FEBS Lett. 313:173-176(1992).
RN   [2]
RP   SEQUENCE OF 106-120 AND 167-184.
RC   STRAIN=CD-1; TISSUE=LUNG, AND LIVER;
RX   MEDLINE; 91378941.
RA   MEDH R.D., SAXENA M., SINGHAL S.S., AHMAD H., AWASTHI Y.C.;
RT   "Characterization of a novel glutathione S-transferase isoenzyme from
RT   mouse lung and liver having structural similarity to rat glutathione
RT   S-transferase 8-8.";
RL   Biochem. J. 278:793-799(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RC   TISSUE=LUNG;
RX   MEDLINE; 98158566.
RA   KRENGEL U., SCHROTER K.H., HOIER H., ARKEMA A., KALK K.H., ZIMNIAK P.,
RA   DIJKSTRA B.W.;
RT   "Crystal structure of a murine alpha-class glutathione S-transferase
RT   involved in cellular defense against oxidative stress.";
RL   FEBS Lett. 422:285-290(1998).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RC   TISSUE=LUNG;
RX   MEDLINE; 99438340.
RA   XIAO B., SINGH S.P., NANDURI B., AWASTHI Y.C., ZIMNIAK P., JI X.;
RT   "Crystal structure of a murine glutathione S-transferase in complex
RT   with a glutathione conjugate of 4-hydroxynon-2-enal in one subunit
RT   and glutathione in the other: evidence of signaling across the dimer
RT   interface.";
RL   Biochemistry 38:11887-11894(1999).
CC   -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC   OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC   -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- MISCELLANEOUS: ON THE BASIS OF IMMUNOLOGICAL AND KINETICS DATA,
CC   GST 5.7 IS DISTINCT FROM ALPHA, MU AND PI CLASSES OF GTS. HOWEVER
CC   IT HAS BEEN POSTULATED THAT THIS PROTEIN MAY BE PART OF A DISTINCT
CC   SUBGROUP WITHIN THIS ALPHA CLASS.
CC   -!- MISCELLANEOUS: THE VARIATIONS WERE FOUND FROM AA SEQUENCING AND
CC   IMPLY THERE ARE MULTIPLE FORMS OF THIS PROTEIN. THESE VARIATIONS
CC   ARE LIKELY TO BE SEX-LINKED AND TISSUE SPECIFIC.
CC   -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
DR   EMBL; L06047; AAA37754.1; -.
DR   PIR; S27234; S27234.
DR   PDB; 1GUK; 08-APR-98.
DR   PDB; 1B48; 29-SEP-99.
DR   MGD; MGI:95857; GSTA.
DR   PFAM; PF00043; GST; 1.
KW   Transferase; Multigene family; Polymorphism; 3D-structure.
FT   VARIANT   115   115   K -> P.
FT   VARIANT   167   167   V -> G.
FT   VARIANT   179   180   PL -> GE.
SQ   SEQUENCE   222 AA; 25576 MW; 715E30C CRC32;

Query Match   51.4%; Score 36; DB 1; Length 222;
Best Local Similarity   60.0%;   Pred. No. 13;
Matches   6; Conservative   6; Mismatches   4; Indels   0; Gaps   0;

Qy   6 IKVIRIVLKY 15
        :|:|||||:|
Db 213 VEVRIVLKF 222

RESULT 6
YKW2_YEAST
ID   YKW2_YEAST   STANDARD;   PRT;   705 AA.
AC   P35995;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   PUTATIVE 82.2 KD TRANSCRIPTIONAL REGULATORY PROTEIN IN FRE2 5'REGION.
GN   YKL222C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94378726.
RA   ALEXANDRAKI D., TZERMIA M.;
RT   "Sequencing of a 13.2 kb segment next to the left telomere of yeast
RT   chromosome XI revealed five open reading frames and recent
RT   recombination events with the right arms of chromosomes III and V.";
RL   Yeast 10:S81-S91(1994).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC   CLUSTER DOMAIN.
DR   EMBL; X75950; CAA53551.1; -.
DR   EMBL; Z28222; CAA82067.1; -.
DR   PIR; S38066; S38066.
DR   PIR; S43548; S43548.
DR   HSSP; P12351; 1PYC.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS00048; ZN2_CY6_FUNGAL_2; 1.
DR   PFAM; PF00172; Zn_clus; 1.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Nuclear protein; Zinc; Metal-binding.
FT   DNA_BIND   24   52   ZN(2)-CYS(6), FUNGAL-TYPE.
SQ   SEQUENCE   705 AA; 82248 MW; 9663DA3B CRC32;

Query Match   51.4%; Score 36; DB 1; Length 705;
Best Local Similarity   46.7%;   Pred. No. 40;
Matches   7; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
        |:||:|: | ||:|
Db 541 AFRALIQIYTIFLQY 555

RESULT 7
SRB8_YEAST
ID   SRB8_YEAST   STANDARD;   PRT;   1427 AA.
AC   P25648;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   SUPPRESSOR OF RNA POLYMERASE B SRB8.
GN   SRB8 OR YCR081W OR YCR81W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]

RP   SEQUENCE FROM N.A.
RX   MEDLINE: 95293223;
RA   HENGARTNER C.J., THOMPSON C.M., ZHANG J., CHAO D.M., LIAO S.M.,
RA   KOLESKE A.J., OKAMURA S., YOUNG R.A.;
RT   "Association of an activator with an RNA polymerase II holoenzyme.";
RL   Genes Dev. 9:897-910(1995).
RN   [2]
RP   REVISIONS
RA   GROMADKA R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 531-1427 FROM N.A.
RA   FELDMANN H., MANNHAUPT G., VETTER I.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE
CC   MEDIATOR OF ACTIVATION SUBCOMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X59720; CAA42268.1; -.
DR   PIR; S19496; S19496.
DR   TRANSFAC; T02152; -.
DR   SGD; L0003007; SRB8.
KW   Nuclear protein.
SQ   SEQUENCE   1427 AA; 166859 MW; 6B732E51 CRC32;

Query Match   51.4%; Score 36; DB 1; Length 1427;
Best Local Similarity   77.8%; Pred. No. 81;
  Matches   7; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   2 YRLLIKVIR 10
              ||||||:||
Db   1200 YHLLIKIIR 1208

RESULT 8
RRPL_SENDE
ID   RRPL_SENDE   STANDARD;   PRT;   2048 AA.
AC   P06829;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Sendai virus (strain Enders).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 86317720.
RA   MORGAN E.M., RAKESTRAW K.M.;
RT   "Sequence of the Sendai virus L gene: open reading frames upstream of
RT   the main coding region suggest that the gene may be polycistronic.";
RL   Virology 154:31-40(1986).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC   FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC   SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC   GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- MISCELLANEOUS: THEY ARE FOUND IN CATALYTIC AMOUNTS (APPROXIMATELY
CC   20 TO 30 COPIES PER SENDAI VIRION) IN VIRAL NUCLEOCAPSIDS.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M14887; AAA69579.1; -.
DR   EMBL; D00053; BAA00036.1; -.
DR   PIR; A24293; ZLNZSE.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2048 AA; 231623 MW; 7FA48C13 CRC32;

Query Match   51.4%; Score 36; DB 1; Length 2048;
Best Local Similarity   53.8%; Pred. No. 1.2e+02;
  Matches   7; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLK 14
              |||:|:|:::|
Db   2003 YRFLTKEIKILMK 2015

RESULT 9
RRPL_SEND5
ID   RRPL_SEND5   STANDARD;   PRT;   2228 AA.
AC   P27566;
DT   01-AUG-1992 (Rel. 23, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Sendai virus (strain Z / host mutants).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MUTANTS TS-F1 AND F1-R;
RX   MEDLINE: 90266486.
RA   MIDDLETON Y., TASHIRO M., THAI T., OH J., SEYMOUR J., PRITZER E.,
RA   KLENK H.D., ROTT R., SETO J.T.;
RT   "Nucleotide sequence analyses of the genes encoding the HN, M, NP, P,
RT   and L proteins of two host range mutants of Sendai virus.";
RL   Virology 176:656-657(1990).
RN   [2]
RP   REVISIONS TO 581 AND 971.
RA   MIDDLETON Y.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MUTANT F1-R / T-5 REVERTANT;
RX   MEDLINE: 91335752.
RA   TASHIRO M., JAMES I., KARRI S., WAHN K., TOBITA K., KLENK H.D.,
RA   ROTT R., SETO J.T.;
RT   "Pneumotropic revertants derived from a pantropic mutant, F1-R, of
RT   Sendai virus.";
RL   Virology 184:227-234(1991).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC   FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC   SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC   GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- MISCELLANEOUS: THEY ARE FOUND IN CATALYTIC AMOUNTS (APPROXIMATELY
CC   20 TO 30 COPIES PER SENDAI VIRION) IN VIRAL NUCLEOCAPSIDS.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

```
CC   ------
DR   EMBL; M30204; AAB06201.1; -.
DR   EMBL; M30203; AAB06289.1; -.
DR   EMBL; M69046; AAB06295.1; -.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
FT   VARIANT     625     625    E -> G (IN MUTANT F1-R / T-5 REVERTANT).
SQ   SEQUENCE   2228 AA;  252748 MW;  E03E1CF1 CRC32;

Query Match             51.4%;  Score 36;  DB 1;  Length 2228;
Best Local Similarity   53.8%;  Pred. No. 1.3e+02;
Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 YRLLIKVIRIVLK 14
           || || |:|::|
Db    2183 YRFLTKEIKILMK 2195

RESULT   10
RRPL_SENDF
ID   RRPL_SENDF     STANDARD;     PRT;   2228 AA.
AC   Q06996;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Sendai virus (strain Fushimi).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92355738.
RA   GIESECKE H., NEUBERT N., OBERMAIER B., DOMDEY H.;
RT   "Rapid sequencing of the Sendai virus 6.8 kb large (L) gene through
RT   primer walking with an automated DNA sequencer.";
RL   J. Virol. Methods 38:47-60(1992).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- MISCELLANEOUS: THEY ARE FOUND IN CATALYTIC AMOUNTS (APPROXIMATELY
CC       20 TO 30 COPIES PER SENDAI VIRION) IN VIRAL NUCLEOCAPSIDS.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------
DR   EMBL; X58886; CAA41690.1; -.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2228 AA;  252770 MW;  7F139FB8 CRC32;

Query Match             51.4%;  Score 36;  DB 1;  Length 2228;
Best Local Similarity   53.8%;  Pred. No. 1.3e+02;
Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 YRLLIKVIRIVLK 14
           || || |:|::|
Db    2183 YRFLTKEIKILMK 2195

RESULT   11
RRPL_SENDZ
ID   RRPL_SENDZ     STANDARD;     PRT;   2228 AA.
AC   P06447;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Sendai virus (strain Z).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86148492.
RA   SHIODA T., IWASAKI K., SHIBUTA H.;
RT   "Determination of the complete nucleotide sequence of the Sendai
RT   virus genome RNA and the predicted amino acid sequences of the F, HN
RT   and L proteins.";
RL   Nucleic Acids Res. 14:1545-1563(1986).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- MISCELLANEOUS: THEY ARE FOUND IN CATALYTIC AMOUNTS (APPROXIMATELY
CC       20 TO 30 COPIES PER SENDAI VIRION) IN VIRAL NUCLEOCAPSIDS.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------
DR   EMBL; X03614; CAA27273.1; -.
DR   PIR; A04120; ZLNZSV.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2228 AA;  252864 MW;  DD9798FD CRC32;

Query Match             51.4%;  Score 36;  DB 1;  Length 2228;
Best Local Similarity   53.8%;  Pred. No. 1.3e+02;
Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 YRLLIKVIRIVLK 14
           || || |:|::|
Db    2183 YRFLTKEIKILMK 2195

RESULT   12
ATPB_MYCLE
ID   ATPB_MYCLE     STANDARD;     PRT;    485 AA.
AC   P45823;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN   ATPD.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SMITH D.R., ROBISON K.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC       SUBUNIT.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
```

CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC --

CC --
DR EMBL; U15186; AAA63108.1; -.
DR HSSP; P07677; 1SKY.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PFAM; PF00006; ATP-synt_ab; 1.
DR PFAM; PF00306; ATP-synt_ab_C; 1.
KW Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW Hydrogen ion transport.
FT NP_BIND   170   177   ATP (POTENTIAL).
SQ SEQUENCE   485 AA; 53034 MW; 07216783 CRC32;

Query Match   50.0%; Score 35; DB 1; Length 485;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy   2 YRLLIKVIRIVLKY 15
        ||: :||||: :|
Db 376 YRVAQEVIRIQRY 389

RESULT 13
ATPB_MYCTU
ID ATPB_MYCTU STANDARD; PRT; 486 AA.
AC Q10593;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34).
GN ATPD OR RV1310 OR MTCY373.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC   GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC   SUBUNIT.
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC   CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC   SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC   HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC --

CC --
DR EMBL; Z73419; CAA97743.1; -.
DR HSSP; P07677; 1SKY.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PFAM; PF00006; ATP-synt_ab; 1.
DR PFAM; PF00306; ATP-synt_ab_C; 1.
KW Hydrolase; ATP synthesis; CF(1); ATP-binding;
KW Hydrogen ion transport.
FT NP_BIND   171   178   ATP (POTENTIAL).
SQ SEQUENCE   486 AA; 53094 MW; A8001B2F CRC32;

Query Match   50.0%; Score 35; DB 1; Length 486;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy   2 YRLLIKVIRIVLKY 15
        ||: :||||: :|
Db 377 YRVAQEVIRIQRY 390

RESULT 14
CAO3_CANTR
ID CAO3_CANTR STANDARD; PRT; 502 AA.
AC P11355;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ACYL-COENZYME A OXIDASE POX4-2 (EC 1.3.3.6) (ACYL-COA OXIDASE)
DE (FRAGMENT).
GN POX4-2.
OS Candida tropicalis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Candidaceae; Candida.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87280361.
RA SMALL G.M., LAZAROW P.B.;
RT "Import of the carboxy-terminal portion of acyl-CoA oxidase into
RT peroxisomes of Candida tropicalis.";
RL J. Cell Biol. 105:247-250(1987).
CC -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC   H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC   FROM 8 TO 18).
CC -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC   SYSTEM.
CC -!- COFACTOR: FAD.
CC -!- SUBUNIT: HOMOOCTAMER.
CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.
DR PIR; A28584; OXCKAX.
KW Oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW Multigene family.
FT NON_TER   1   1
SQ SEQUENCE   502 AA; 55528 MW; EACE80C4 CRC32;

Query Match   50.0%; Score 35; DB 1; Length 502;
Best Local Similarity 40.0%; Pred. No. 43;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy   1 AYRLLIKVIRIVLKY 15
        :||:| |: | ||:|
Db 100 SYRMLARMSTIALRY 114

RESULT 15
CAO2_CANTR
ID CAO2_CANTR STANDARD; PRT; 708 AA.
AC P06598;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ACYL-COENZYME A OXIDASE II (EC 1.3.3.6) (ACYL-COA OXIDASE) (PXP-4).

Db   306 SYRMLARMSTIALRY 320

Search completed: February 8, 2000, 01:25:56
Job time: 1556 sec

GN   AOX OR POX-4.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Candidaceae; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20336 / PK233;
RX   MEDLINE; 87248070.
RA   MURRAY W.W., RACHUBINSKI R.A.;
RT   "The primary structure of a peroxisomal fatty acyl-CoA oxidase from
RT   the yeast Candida tropicalis pK233.";
RL   Gene 51:119-128(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20336 / PK233;
RX   MEDLINE; 86149279.
RA   OKAZAKI K., TAKECHI T., KAMBARA N., FUKUI S., KUBOTA I., KAMIRYO T.;
RT   "Two acyl-coenzyme A oxidases in peroxisomes of the yeast Candida
RT   tropicalis: primary structures deduced from genomic DNA sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1232-1236(1986).
RN   [3]
RP   SEQUENCE OF 208-709 FROM N.A.
RC   STRAIN=RR1;
RX   MEDLINE; 87280361.
RA   SMALL G.M., LAZAROW P.B.;
RT   "Import of the carboxy-terminal portion of acyl-CoA oxidase into
RT   peroxisomes of Candida tropicalis.";
RL   J. Cell Biol. 105:247-250(1987).
CC   -!- CATALYTIC ACTIVITY: ACYL-COA + O(2) = TRANS-2,3-DEHYDROACYL-COA +
CC   H(2)O(2).(ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC   FROM 8 TO 18).
CC   -!- COFACTOR: FAD.
CC   -!- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC   SYSTEM.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M16193; AAA34322.1; -.
DR   EMBL; M12160; AAA34362.1; -.
DR   EMBL; Y00623; CAA68660.1; -.
DR   EMBL; Y00623; CAA68661.1; ALT_INIT.
DR   EMBL; Y00623; CAA68662.1; ALT_INIT.
DR   PIR; A25123; OXCKX4.
DR   PIR; A29047; OXCKX.
KW   oxidoreductase; Fatty acid metabolism; Peroxisome; Flavoprotein; FAD;
KW   Multigene family.
FT   INIT_MET   0   0
FT   CONFLICT   216   216   Q -> E (IN REF. 3).
FT   CONFLICT   245   245   P -> A (IN REF. 2 AND 3).
FT   CONFLICT   335   335   N -> K (IN REF. 2).
FT   CONFLICT   359   393   FRAMESHIFT ERROR (IN REF. 2 AND 3).
FT   CONFLICT   435   435   H -> Y (IN REF. 2).
FT   CONFLICT   462   462   G -> A (IN REF. 2).
FT   CONFLICT   495   495   E -> S (IN REF. 2).
FT   CONFLICT   576   578   ELA -> DLV (IN REF. 2).
FT   CONFLICT   697   697   Q -> E (IN REF. 2 AND 3).
SQ   SEQUENCE   708 AA; 79041 MW; D97A4EC8 CRC32;

Query Match   50.0%; Score 35; DB 1; Length 708;
Best Local Similarity 40.0%; Pred. No. 61;
Matches   6; Conservative   5; Mismatches   4; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
        :||:||::: | |:||

```
            GenCore version 4.5
      Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 8, 2000, 13:17:53 ; Search time 176.54 Seconds
                                         (without alignments)
                                         5.891 Million cell updates/sec

Title:            US-08-653-294-21
Perfect score:    70
Sequence:         1 AYRLLIKVIRIVLKY 15

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters:   225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing:  Minimum Match 0%
                  Listing first 45 summaries

Database :   SPTREMBL_12:*
          1: sp_archea:*
          2: sp_bacteria:*
          3: sp_fungi:*
          4: sp_human:*
          5: sp_invertebrate:*
          6: sp_mammal:*
          7: sp_mhc:*
          8: sp_organelle:*
          9: sp_phage:*
         10: sp_plant:*
         11: sp_rodent:*
         12: sp_virus:*
         13: sp_vertebrate:*
         14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	57.1	2810	5	Q20456	Q20456 caenorhabdi
2	39.5	56.4	746	3	Q08754	Q08754 saccharomyc
3	38	54.3	771	5	Q20026	Q20026 caenorhabdi
4	37	52.9	98	2	Q45637	Q45637 bacillus sp
5	37	52.9	1696	5	Q21436	Q21436 caenorhabdi
6	36	51.4	100	2	Q69234	Q69234 bacillus ce
7	36	51.4	131	2	Q66643	Q66643 aquifex aeo
8	36	51.4	141	4	Q75379	Q75379 homo sapien
9	36	51.4	186	5	Q96287	Q96287 plasmodium
10	36	51.4	187	2	Q84836	Q84836 chlamydia t
11	36	51.4	282	1	Q30154	Q30154 archaeoglob
12	36	51.4	284	2	Q50343	Q50343 lactobacill
13	36	51.4	325	2	Q9z9x8	Q9Z9X8 frateuria s
14	36	51.4	381	2	Q86996	Q86996 acinetobact
15	36	51.4	492	3	Q08966	Q08966 saccharomyc
16	36	51.4	1980	12	Q84185	Q84185 human paral
17	36	51.4	2018	12	Q20487	Q20487 caenorhabdi
18	36	51.4	2223	12	Q9WFZ5	Q9WFZ5 human parai
19	36	51.4	2228	12	Q98705	Q98705 sendai viru
20	35.5	50.7	1036	10	O49323	O49323 arabidopsis
21	35	50.0	83	12	Q89074	Q89074 variola vir
22	35	50.0	175	2	Q17265	Q17265 brugia paha
23	35	50.0	240	2	Q68993	Q68993 chlorobium
24	35	50.0	341	1	Q9YDV4	Q9YDV4 aeropyrum p
25	35	50.0	555	5	Q76770	Q76770 dictyosteli
26	35	50.0	791	10	Q80962	Q80962 arabidopsis
27	35	50.0	1687	11	Q9WUF6	Q9WUF6 mus musculu
28	35	50.0	1740	5	Q23391	Q23391 dictyosteli
29	35	50.0	1740	8	Q9XPI9	Q9XPI9 dictyosteli
30	35	50.0	1782	13	O42180	O42180 fugu rubrip
31	35	50.0	1784	4	O75275	O75275 homo sapien
32	35	50.0	1786	5	Q17344	Q17344 caenorhabdi
33	35	50.0	1809	5	Q17487	Q17487 caenorhabdi
34	35	50.0	1814	11	Q61037	Q61037 mus musculu
35	35	50.0	1815	5	Q17488	Q17488 caenorhabdi
36	35	50.0	1827	5	Q97275	Q97275 plasmodium
37	35	50.0	1867	5	Q17486	Q17486 caenorhabdi
38	35	50.0	2039	5	Q17489	Q17489 caenorhabdi
39	35	50.0	2228	12	O55528	O55528 sendai viru
40	35	50.0	2228	12	O55530	O55530 sendai viru
41	35	50.0	3973	5	O96204	O96204 plasmodium
42	35	50.0	6994	5	Q17343	Q17343 caenorhabdi
43	34	48.6	180	5	Q9YC59	Q9YC59 aeropyrum p
44	34	48.6	185	2	Q9Z973	Q9Z973 chlamydia p
45	34	48.6	208	8	O21527	O21527 clethrionom

ALIGNMENTS

```
RESULT 1
Q20456

ID   Q20456        PRELIMINARY;       PRT;   2810 AA.
AC   Q20456;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HUM-4 PROTEIN.
GN   HUM-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   COTTAGE A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z66563; CAA91469.1; -.
DR   PFAM; PF00612; IQ; 2.
DR   PFAM; PF00063; myosin_head; 4.
DR   PFAM; PF00784; MYTH4; 2.
SQ   SEQUENCE   2810 AA;  323526 MW;  62742B6C CRC32;

Query Match             57.1%;  Score 40;  DB 5;  Length 2810;
Best Local Similarity   35.7%;  Pred. No. 1.6e+02;
Matches    5;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;
```

Qy   2 YRLLIKVIRIVLKY 15
          :||:|| ::|:|||
Db   1102 FRLSVEIFKLILKY 1115

RESULT 2
ID   Q08754   PRELIMINARY;   PRT;   746 AA.
AC   Q08754;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   ORF YOR299W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CZIEPLUCH C., JAUNIAUX J.C., KORDES E., POIREY R., PUJOL A.,
RA   TOBIASCH E.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z75207; CAA99528.1; -.
SQ   SEQUENCE   746 AA; 84829 MW; 01352AC2 CRC32;

Query Match
Best Local Similarity 56.4%;   Score 39.5; DB 3; Length 746;
Matches 9;   Conservative 45.0%;   Pred. No. 56;
Mismatches 5;   Indels 1;   Gaps 1;

Qy   1 AYRLLIKVIRI----VLKY 15
          |||| ::::| :||
Db   502 AYRLLTEIVQITGWENLLKY 521

RESULT 3
ID   Q20026   PRELIMINARY;   PRT;   771 AA.
AC   Q20026;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CODED FOR BY C. ELEGANS CDNA CEESB82F.
GN   F35C8.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   WU X.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;

RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U40941; AAA81709.1; -.
SQ   SEQUENCE   771 AA; 87310 MW; 5AE2EE3F CRC32;

Query Match
Best Local Similarity 54.3%;   Score 38; DB 5; Length 771;
Matches 5;   Conservative 45.5%;   Pred. No. 1.1e+02;
Mismatches 6;   Indels 0;   Gaps 0;

Qy   5 LIKVIRIVLKY 15
          ::||||:||:|:|
Db   578 IVKIIRVLLEY 588

RESULT 4
ID   Q45537   PRELIMINARY;   PRT;   98 AA.
AC   Q45537;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 10.9 KD PROTEIN.
OS   Bacillus sp., Bacillus megaterium, and Exiguobacterium sp.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RC607;
RX   MEDLINE; 89123092.
RA   WANG Y., MOORE M., LEVINSON H.S., SILVER S., WALSH C., MAHLER I.;
RT   "Nucleotide sequence of a chromosomal mercury resistance determinant
RT   from a Bacillus sp. with broad-spectrum mercury resistance.";
RL   J. Bacteriol. 171:83-92(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MK64-1;
RA   MINAKHIN L.S.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MK64-1, TC38-2B;
RX   MEDLINE; 98195721.
RA   BOGDANOVA E.S., BASS I.A., MINHAKHIN L.S., PETROVA M.A., MINDLIN S.Z.,
RA   VOLODIN A.A., KALYAEVA E.S., TIEDGE G.M., HOBMAN J.L., BROWN N.L.,
RA   NIKIFIROV V.G.;
RT   "Horizontal spread of mer operons among gram-positive bacteria in
RT   natural environments.";
RL   Microbiology 144:609-620(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TC38-2B;
RA   MINAKHIN L.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TC38-2B;
RA   MINAKHIN L.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF138877; AAA83975.1; -.
DR   EMBL; Y09907; CAA71043.1; -.
DR   EMBL; X99457; CAA67820.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   98 AA; 10900 MW; F1D67BC1 CRC32;

Query Match
Best Local Similarity 52.9%;   Score 37; DB 2; Length 98;
Matches 7;   Conservative 58.3%;   Pred. No. 22;
Mismatches 2;   Indels 3;   Gaps 0;

Qy   2 YRLLIKVIRIVL 13
          |||| ||| |||
Db   46 YRLLFSIVTIVL 57

RESULT 5
Q21436
ID   Q21436   PRELIMINARY;   PRT;   1696 AA.
AC   Q21436;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   K12D12.2 PROTEIN.
GN   K12D12.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   COLES L.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L. WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z49069; CAA88864.1; -.
SQ   SEQUENCE 1696 AA; 191324 MW; 6EDE7750 CRC32;

Query Match   52.9%; Score 37; DB 5; Length 1696;
Best Local Similarity   53.3%;
Matches   8; Conservative   4; Mismatches   3; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
          |:||||.:|.:||:
Db 121 AHRLLIATMRQMLKW 135

RESULT 6
069234
ID   069234   PRELIMINARY;   PRT;   100 AA.
AC   069234;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRANSPORT PROTEIN.
OS   Bacillus cereus.
OG   Plasmid pKLH302, and plasmid pKLH301.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   PLASMID=PKLH302; TRANSPOSON=TN1546-LIKE, AND TN3-TYPE;
RA   MINAKHIN L.S.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   PLASMID=PKLH301; TRANSPOSON=TN1546-LIKE, AND TN3-TYPE;
RX   MEDLINE; 98195721.
RA   BOGDANOVA E.S., BASS I.A., MINHAKHIN L.S., PETROVA M.A., MINDLIN S.Z.,
RA   VOLODIN A.A., KALYAEVA E.S., TIEDGE G.M., HOBMAN J.L., BROWN N.L.,
RA   NIKIFIROV V.G.;
RT   "Horizontal spread of mer operons among gram-positive bacteria in
RT   natural environments.";
RL   Microbiology 144:609-620(1998).
DR   EMBL; Y09024; CAA70222.1; -.
DR   EMBL; Y09027; CAA70244.1; -.
KW   Plasmid.
SQ   SEQUENCE 100 AA; 11126 MW; 381D4576 CRC32;

Query Match   51.4%; Score 36; DB 2; Length 100;
Best Local Similarity   50.0%; Pred. No. 34;
Matches   6; Conservative   3; Mismatches   3; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVL 13
          |||| .:|:||
Db   46 YRLLFSIVTIIL 57

RESULT 7
066643
ID   066643   PRELIMINARY;   PRT;   131 AA.
AC   066643;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 15.4 KD PROTEIN.
GN   AQ_293.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000682; AAC06601.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 131 AA; 15393 MW; E36D7DD3 CRC32;

Query Match   51.4%; Score 36; DB 2; Length 131;
Best Local Similarity   28.6%; Pred. No. 44;
Matches   4; Conservative   8; Mismatches   2; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLKY 15
          :::|:||.:::|
Db   46 HKILVKVLNVIIIY 59

RESULT 8
075379
ID   075379   PRELIMINARY;   PRT;   141 AA.
AC   075379;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   VAMP 4.
GN   VAMP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98221163.
RA   ADVANI R.J., BAE H.R., BOCK J.B., CHAO D.S., DOUNG Y.C., PREKERIS R.,

RA YOO J.S., SCHELLER R.H.;
RT "Seven novel mammalian SNARE proteins localize to distinct membrane
RT compartments.";
RL J. Biol. Chem. 273:10317-10324(1998).
DR EMBL; AF044310; AAC24032.1; -.
DR PFAM; PF00957; synaptobrevin; 1.
DR PRINTS; PR00219; SYNAPTOBREVN.
SQ SEQUENCE 141 AA; 16366 MW; 3C737E45 CRC32;

Query Match   51.4%; Score 36; DB 4; Length 141;
Best Local Similarity   53.3%; Pred. No. 47;
Matches   8; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
          | |: || :|| |:||
Db   125 AILLIVIILIVMKY 139

RESULT 9
ID 096287   PRELIMINARY;   PRT;   186 AA.
AC 096287;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PREDICTED INTEGRAL MEMBRANE PROTEIN.
GN PFB0985C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99021743.
RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001431; AAC71985.1; -.
SQ SEQUENCE 186 AA; 21961 MW; B165839F CRC32;

Query Match   51.4%; Score 36; DB 5; Length 186;
Best Local Similarity   53.3%; Pred. No. 61;
Matches   8; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   1 AYRLLIKVIRIVLKY 15
          |:||| :|| :|| :|
Db   145 AFILLIFIIHIVARY 159

RESULT 10
ID 084836   PRELIMINARY;   PRT;   187 AA.
AC 084836;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE RRNA METHYLASE (POSSIBLE).
GN YGGH.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";

RL Science 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE001356; AAC68426.1; -.
KW Methyltransferase.
SQ SEQUENCE 187 AA; 21768 MW; 9AF9D1A5 CRC32;

Query Match   51.4%; Score 36; DB 2; Length 187;
Best Local Similarity   42.9%; Pred. No. 61;
Matches   6; Conservative   3; Mismatches   5; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLKY 15
          | |: | :||||
Db   130 YNYLVNAITVMLKY 143

RESULT 11
ID 030154   PRELIMINARY;   PRT;   282 AA.
AC 030154;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE HYPOTHETICAL 33.0 KD PROTEIN.
GN AF0082.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 /DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001100; AAB91148.1; -.
DR TIGR; AF0082;
KW Hypothetical protein.
SQ SEQUENCE 282 AA; 33040 MW; 2712078C CRC32;

Query Match   51.4%; Score 36; DB 1; Length 282;
Best Local Similarity   53.8%; Pred. No. 90;
Matches   7; Conservative   4; Mismatches   2; Indels   0; Gaps   0;

Qy   2 YRLLIKVIRIVLK 14
          ||::::||:|| ||
Db   269 YRHLMRLIRMTLK 281

RESULT 12
ID 050343   PRELIMINARY;   PRT;   284 AA.
AC 050343;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

```
DT   01-JUN-1998 (TrEMBLrel. 05, Last annotation update)
DE   HYPOTHETICAL 33.6 KD PROTEIN.
OS   Lactobacillus helveticus.
OG   Plasmid pLH1.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Lactobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15009;
RA   THOMPSON K., MCCONVILLE K.J., MCREYNOLDS C., FOLEY S.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ222725; CAA10961.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   284 AA;  33632 MW;  E77494B0 CRC32;

Query Match             51.4%;
Best Local Similarity   28.6%;        Score 36;  DB 2;  Length 284;
                                      Pred. No. 91;
Matches   4;  Conservative   8;  Mismatches   1;  Indels   2;  Gaps   0;

Qy     2 YRLLIKVIRIVLKY 15
         |::::|:::::::||
Db   231 YKSVVKFLKLLIKY 244

RESULT 13
Q929X8
ID   Q929X8      PRELIMINARY;      PRT;   325 AA.
AC   Q929X8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ELECTRON TRANSFER PROTEIN.
OS   Frateuria sp. ANA-18.
OC   Bacteria; Proteobacteria; gamma subdivision; Frateuria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANA-18;
RX   MEDLINE; 99132292.
RA   MURAKAMI S., TAKASHIMA A., TAKEMOTO J., TAKENAKA S., SHINKE R.,
RA   AOKI K.;
RT   "Cloning and sequence analysis of two catechol-degrading gene clusters
RT   from the aniline-assimilating bacterium Frateuria species ANA-18.";
RL   Gene 226:189-198(1999).
DR   EMBL; AB009373; BAA75213.1; -.
DR   HSSP; P33164; 2PIA.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   Iron-sulfur.
SQ   SEQUENCE   325 AA;  34987 MW;  6A35946C CRC32;

Query Match             51.4%;
Best Local Similarity   60.0%;        Score 36;  DB 2;  Length 325;
                                      Pred. No. 1e+02;
Matches   9;  Conservative   1;  Mismatches   1;  Indels   5;  Gaps   0;

Qy     1 AYRLLIKVIRIVLKY 15
         |||||  | |||
Db   127 AYRLLKQEQRFVLHY 141

RESULT 14
O86996
ID   O86996      PRELIMINARY;      PRT;   381 AA.
AC   O86996;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 41.4 KD PROTEIN.
OS   Acinetobacter lwoffii K24.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Moraxellaceae; Acinetobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K24;
RA   KIM S.I., LEEM S.-H., CHOI J.S., CHUNG Y.H., KIM S., PARK Y.-M.,
RA   HA K.-S.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77659; AAC31770.1; -.
DR   HSSP; P33164; 2PIA.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PFAM; PF00111; fer2; 1.
KW   Hypothetical protein; Iron-sulfur.
SQ   SEQUENCE   381 AA;  41377 MW;  BE9CD1C2 CRC32;

Query Match             51.4%;
Best Local Similarity   60.0%;        Score 36;  DB 2;  Length 381;
                                      Pred. No. 1.2e+02;
Matches   9;  Conservative   1;  Mismatches   5;  Indels   5;  Gaps   0;

Qy     1 AYRLLIKVIRIVLKY 15
         |||||  : |||
Db   183 AYRLLKQEQRFVLHY 197

RESULT 15
Q08966
ID   Q08966      PRELIMINARY;      PRT;   492 AA.
AC   Q08966;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   CHROMOSOME XVI READING FRAME ORF YPL219W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RIEGER M., MUELLER-AUER S., SCHAEFER M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73575; CAA97934.1; -.
SQ   SEQUENCE   492 AA;  55430 MW;  6C549DE0 CRC32;

Query Match             51.4%;
Best Local Similarity   50.0%;        Score 36;  DB 3;  Length 492;
                                      Pred. No. 1.5e+02;
Matches   7;  Conservative   3;  Mismatches   4;  Indels   4;  Gaps   0;

Qy     1 AYRLLIKVIRIVLK 14
         |:|::| ||| ||
Db   416 AHRIIISTIRIATK 429

Search completed: February 8, 2000, 19:16:09
Job time: 21496 sec
```

```
OM of: US-08-653-294-21 to: GenEmbl:*      out_format : pfs
Date: Feb 8, 2000 10:21 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV-xlp
 -Q=/cgn1_1/USPTO_spool/US08653294/runat_04022000_160701_15779/app_query.fasta.2
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US08653294
 -NCPU=3 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-653-294-21
Query length: 15
Database: GenEmbl:*
Database sequences: 821193
Database length: -1518192014
Search time (sec): 10176.920000
```

```
gb_htg7:AC012083                      -      43.00     83.83   1.2e+04     222153  ! AC012083 Homo sapiens clo
gb_pri:HSH2F32                        -      42.00    120.90    103.64       1812  ! X78926 H.sapiens H2F3 mRNA
gb_pat:E06738                         +      42.00    119.06    131.24       2255  ! E06738 Rhizopus niveus chit

seq_name: gb_pl3:AF163820

seq_documentation_block:
LOCUS       AF163820       3000 bp    DNA            PLN      01-DEC-1999
DEFINITION  Arabidopsis thaliana endoxyloglucan transferase (XTR2) gene,
            complete cds.
ACCESSION   AF163820
VERSION     AF163820.1   GI:5533310
KEYWORDS    .
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 3000)
AUTHORS     Akamatsu,T., Hanzawa,Y., Ohtake,Y., Takahashi,T., Nishitani,K. and
            Komeda,Y.
TITLE       Expression of endoxyloglucan transferase genes in acaulis mutants
            of Arabidopsis
JOURNAL     Plant Physiol. 121 (3), 715-721 (1999)
REFERENCE   2  (bases 1 to 3000)
AUTHORS     Takahashi,T.
TITLE       Direct Submission
JOURNAL     Submitted (30-JUN-1999) Division of Biological Sciences, Graduate
            School of Science, Hokkaido University, N10, W8, Sapporo 060-0810,
            Japan

FEATURES             Location/Qualifiers
     source          1..3000
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia-0"
                     /db_xref="taxon:3702"
     mRNA            join(<1118..1301,1389..1489,1594..1802,2276..>2780)
                     /gene="XTR2"
                     /product="endoxyloglucan transferase"
     gene            <1118..>2780
                     /gene="XTR2"
     CDS             join(1118..1301,1389..1489,1594..1802,2276..2780)
                     /gene="XTR2"
                     /codon_start=1
                     /product="endoxyloglucan transferase"
                     /protein_id="AAD45124.1"
                     /db_xref="GI:5533311"
                     /translation="MGFITRFLVFEMSLFTSLVSGFALQKLPLIQFDEGYTQLFGDQNL
                     IVHRDGKSVRLTLDERTGSGFVSNDIYLHGFFSSSIKLPADYSAGVVIAFYLSNGDLY
                     EKNHDEIDFEFLGNIRGREWRIQTNIYGNGSTHLGREERNLWFPTEDFHQYSLWS
                     LSHIFFYDNVPIREVKRTASMGDFPAKPMSLYSTIWDGSKWATDGKYGVNYKYAP
                     YVSQFTDLIIHGCAVDPTEKFPSCKDEAVQNLRLASEITESQRNKMEIFRQKHMTYSY
                     CYDRMRYKVVLSECVVNPAEAKRLRVYDPVTFGGIPHGHRRGKHRSRSRLARTESI"
BASE COUNT    906 a    527 c    516 g   1048 t    3 others
ORIGIN

alignment_scores:
         Quality:  48.00              Length:       15
         Ratio:    4.000              Gaps:          0
Percent Similarity:  80.000    Percent Identity: 66.667

alignment_block:
US-08-653-294-21 x AF163820/rev  ..

Align seq 1/1 to reverse of: AF163820  from: 1  to: 3000

       1 AlaTyrArgLeuLeuLysValIleArglleValIleuLysTyr 15
         |||||T|||||:::  ||||||||  |||||||:||||::||||
     650 GCATATCGACTGATAGCTAAGGTTATTACAATTTGTTAGCATAT 606

seq_name: gb_inl:CEF56G4
```

```
score_list:
sequence                    Strd  Orig   ZScore   EScore    Len  ! Documentation ..
gb_pl3:AF163820              +    48.00  137.09   12.99     3000  ! AF163820 Arabidopsis thaliana
gb_inl:CEF56G4              +    48.00  115.71   201.69    38062  ! Z81552 Caenorhabditis elegans
gb_htg1:CEY53H1_2          +    48.00  106.77   634.22   110000  ! Continuation (3 of 4) of CEY
gb_pl2:AC006917            -    48.00  105.19   776.59   132699  ! Continuation (3 of 4) of CEY
gb_htg5:AC016012           -    47.00  107.33   590.34    68676  ! AC016012 Homo sapiens clone R
gb_pl1:AB017067            +    46.00  102.26   1.1e+03   83689  ! AB017067 Arabidopsis thaliana
gb_ba2:AF027868            -    46.00  101.89   1.2e+03   87500  ! AF027868 Bacillus subtilis ch
gb_pl2:ATF13C5             -    46.00   99.29   1.7e+03  119111  ! AL021711 Arabidopsis thalian
gb_htg6:AC011694           +    46.00   96.78   2.3e+03  160557  ! AC011694 Homo sapiens clone
gb_bal:BSUB0011            -    46.00   94.61   3.0e+03  207730  ! 299114 Bacillus subtilis com
gb_pl1:NIMIDRPM            +    45.00  117.97  150.93     8642  ! X52106 Neurospora intermedia m
gb_pl1:SC957IX             +    45.00  107.67  565.24    29366  ! Z49810 S.cerevisiae chromosom
gb_htg5:AC015353           -    45.00  107.64  567.17    29459  ! AC015353 Drosophila melanogas
gb_pl1:SPAC9G1             +    45.00  107.22  598.95    30985  ! 298763 S.pombe chromosome I c
gb_inl:CELF29G9            -    45.00  104.51  847.82    42751  ! AF016440 Caenorhabditis elega
gb_htg2:HS758N20           -    45.00   94.66   3.0e+03  137884  ! AL035705 Homo sapiens chromo
gb_htg5:AC014946           -    44.00   99.64   1.6e+03   50885  ! AC014946 Drosophila melanogas
gb_pl1:ATF5C23             +    44.00   95.11   2.8e+03   87104  ! AL049500 Arabidopsis thaliana
gb_htg3:AC009916           +    44.00   93.32   3.6e+03  107727  ! AC009916 Drosophila melanogas
gb_pr4:AC004997            +    44.00   93.21   3.6e+03  109238  ! AC004997 Homo sapiens PAC cl
gb_pr3:HS394F12            -    44.00   93.12   3.7e+03  109279  ! Z83823 Human DNA sequence fr
gb_htg4:AC010487           -    44.00   91.32   4.6e+03  110409  ! AC010487 Homo sapiens chromo
gb_htg4:AC010014           +    44.00   91.32   4.6e+03  136702  ! AC010014 Drosophila melanoga
gb_pr3:AC005585            +    44.00   90.00   5.1e+03  158257  ! AC005585 Homo sapiens chromo
gb_pr2:AC006802            +    44.00   83.87   1.2e+04  331326  ! AC006802 Caenorhabditis eleg
gb_ro:AB013484S8           +    43.00  113.98  251.68     6178  ! AB013491 Mus musculus mSim1 ge
gb_inl:CELH42K12           +    43.00  107.04  613.20    14097  ! AF026207 Caenorhabditis elega
gb_htg6:AC016136           +    43.00  100.28   1.5e+03   31460  ! AC016136 Homo sapiens clone R
gb_inl:CER07H5             -    43.00   99.67   1.6e+03   33802  ! Z81107 Caenorhabditis elegans
gb_inl:CEC48D1             +    43.00   98.28   1.9e+03   39908  ! Z81049 Caenorhabditis elegans
gb_inl:CELC52B11           -    43.00   98.04   1.9e+03   41039  ! U41276 Caenorhabditis elegans
gb_pr3:HS300I2             -    43.00   94.33   3.1e+03   63796  ! AL035660 Human DNA sequence f
gb_htg2:AC005636           +    43.00   93.00   3.7e+03   74686  ! AC005636 Drosophila melanogas
gb_htg2:AC008181           +    43.00   92.15   4.1e+03   82632  ! AC008181 Drosophila melanogas
gb_pl1:AB024038            +    43.00   92.03   4.2e+03   83859  ! AB024038 Arabidopsis thaliana
gb_htg3:AC009979           +    43.00   91.94   4.2e+03   84719  ! AC009979 Homo sapiens, *** SE
gb_htg5:AC013247           -    43.00   91.29   4.6e+03   91503  ! AC013247 Drosophila melanogas
gb_htg5:AC011678           -    43.00   86.14   8.9e+03  168742  ! AC011678 Homo sapiens clone
gb_htg7:AC017008           -    43.00   85.98   9.1e+03  172065  ! AC017008 Homo sapiens clone
gb_htg2:AC008163           -    43.00   85.73   9.4e+03  177261  ! AC008163 Homo sapiens clone
gb_pr4:AC007052            +    43.00   85.61   9.6e+03  179726  ! AC007052 Homo sapiens chromo
gb_htg4:AC012010           -    43.00   85.42   9.8e+03  183357  ! AC012010 Homo sapiens chromo
```

```
seq_documentation_block:
LOCUS       CEF56G4     38062 bp    DNA         INV       08-OCT-1999
DEFINITION  Caenorhabditis elegans cosmid F56G4, complete sequence.
ACCESSION   Z81552
VERSION     Z81552.1  GI:3217528
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 38062)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
REFERENCE   2  (bases 1 to 38062)
AUTHORS     Lennard,N.
TITLE       Direct Submission
JOURNAL     Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 13, 1998 this sequence version replaced gi:1665968.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormaceclass=Sequence&object=F56G4
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            This sequence is the entire insert of clone F56G4. The true right
            end of clone M04D5 is at 1937 in this sequence. The start of this
            sequence (1..100) overlaps with the end of sequence Z83118.
            The end of this sequence (37961..38062) overlaps with the start of
            sequence AL117201.
FEATURES             Location/Qualifiers
     source          1..38062
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="F56G4"
     gene            9882..11016
                     /gene="pes-2"
     CDS             join(9882..9955,10019..10840,10917..11016)
                     /gene="pes-2"
                     /note="predicted using Genefinder; cDNA EST yk254h12.3
                     comes from this gene; cDNA EST yk427e7.3 comes from this
                     gene; cDNA EST yk427e7.5 comes from this gene; cDNA EST
                     yk455a6.3 comes from this gene; cDNA EST yk455a6.5 comes
                     from this gene; cDNA EST yk465b3.3 comes from this gene;
                     cDNA EST yk465b3.5 comes from this gene; cDNA EST
                     EMBL:T00466 comes from this gene; cDNA EST EMBL:T00835
                     comes from this gene; cDNA EST yk404g9.3 comes from this
                     gene; cDNA EST yk404g9.5 comes from this gene; cDNA EST
                     yk522g11.3 comes from this gene; cDNA EST yk621e3.3 comes
                     from this gene; cDNA EST yk630h10.3 comes from this gene;
                     cDNA EST yk635a12.3 comes from this gene; cDNA EST
                     yk667d4.3 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB04484.1"
                     /db_xref="GI:3877756"
                     /translation="MPVKVLNFPSVVLQHFLVGLTSTELFELTQCSLKSKDRVRPYEK
                     SNKTFRMAVDFQRNWVDIRGVYRFVVKELVDGKLPKTYGKMGVREFGDKVQVELDEN
                     EKTINIFWPTEKEDRTFAVMKVAEHFIYTLGLGEFESVTLGQENHAGRIIDWANEKN
                     LNIANFTIGNVDKEHDDVDKVAEKLYSASFFKKVSANFKCLINPTEAFKPNEFLPGIP
                     KNTNMIFFHGAHWMTIDTLKDFTACSVILYDCKFTNKDINEFIKLWKNKEFERIFYIS
                     INVPQGSGFKLTEVIAGLDGFGEIENDATTMKREDGKEIEVSVLDEKKQIRIVVD"
     gene            complement(12813..13947)
                     /gene="F56G4.3"
     CDS             complement(join(12813..12912,12989..13810,13874..13947))
                     /gene="F56G4.3"
                     /codon_start=1
                     /protein_id="CAB04485.1"
                     /db_xref="GI:3877757"
                     /db_xref="SPTREMBL:Q27347"
                     /translation="MPVKVLNFPSVVLQHFLVGLTSTELFELTQCSLKSKDRVRPYEK
                     SNKTFRMAVDFQRNWVDIRGVYRFVVKELVDGKLPKTYGKMGVREFGDKVQVELDEN
                     EKTINIFWPTEKEDRTFAVMKVAEHFIYTLGLGEFESVTLGQENHAGRIIDWANEKN
                     LNIANFTIGNVDKEHDDVDKVAEKLYSASFFKKVSANFKCLINPTEAFKPNEFLPGIP
                     KNTNMIFFHGAHWMTIDTLKDFTACSVILYDCKFTNKDINEFIKLWKNKEFERIFYIS
                     INVPQGSGFKLTEVIAGLDGFGEIENDATTMKREDGKEIEVSVLDEKKQIRIVVD"
     gene            complement(19032..28678)
                     /gene="F56G4.6"
     CDS             complement(join(19032..19201,19777..19892,21407..21732,
                     22710..22888,22991..23276,24996..25192,26681..27327,
                     28360..28469,28562..28678))
                     /gene="F56G4.6"
                     /codon_start=1
                     /protein_id="CAB04488.1"
                     /db_xref="GI:3877760"
                     /translation="MVYNVLLSLPMMIQARWELLALLGSGLIKQMDEVIFKEIMACA
                     SDQDLAARCSSVMLQISSNSDTCDKTAAIIFKGLJTSPVAAVRSNVHRSWLCRIITLKN
                     SQKCVETLKLCKKSIKTGNCSKILPISFRELPGLTCVDLQWRAILDEEPGDFSAFC
                     YFSVLLSLKNASSQVHIQIPANDEVLAELRNSGSLDVLKVCSKYKIYPEPLENWILE
                     HFSPEMMEYIDFSMLKTFSAGFSQVFEEKKLEICQEKLEKFYPAEDLGRLNVHVDLSK
                     EKLANETLEQFLQIHESEEYPSVAGVMENLSKONIDERLFCQKICLLINDIMHEKYSI
                     LKEIRREMTEFIFKIITRIPLRSVVDYVGATLKTEALQNPDFKNQIYQMSCQLLRLD
                     KENARPLSLYHCFIGTCQQDTTARISEIWAIIRGLKDFKTMEMLAKVVTSSEI
                     QIDPSLLSDMTICSLGLHPTRFNTESTQLFVASISKLLDFRNESERIMLFQATANTPT
                     LMDLISSILSRPSTDFSISGFQMILYFTRFHIGSMEFYPYNHLLLITKIINFCHWSI
                     RKFAKEPMILSISVDAICGFLPWSYLSEILNGLEVENDPVEQLIAQNLKHRNFCVLL
                     AEHDKQDTLTIEDFEKLPKSEFIVEILRKDAGNGQKWKNQYPPRGFLDNLMESFAA
                     TNSCINPEEIQNWLEKCLGMEIVRKKIIDFY"
     gene            29270..34711
                     /gene="F56G4.4"
     CDS             join(29270..29320,29514..29916,31073..31341,32182..32473,
                     33777..34108,34646..34711)
                     /gene="F56G4.4"
                     /note="cDNA EST yk363cl1.5 comes from this gene; cDNA EST
                     yk607el.3 comes from this gene; cDNA EST yk672a5.3 comes
                     from this gene"
                     /codon_start=1
                     /protein_id="CAB04486.1"
                     /db_xref="GI:3877758"
                     /translation="MAEQWVSTGKRFCDICKVWFGNNRASQDHDRGERHKAMLQQRI
                     RETMQKGKKQELADMKLNGTLAKMEAAAASMARNGEAIVAGPSLPATGLRMNIFDPS
                     KHKDVGSMAKEMSSRREAAGNVKRGAQSAAVGAPNVKIITKDDVVSAYRDLTNPNAAGP
                     ALPEHLAVGYMPETQVCEIQWVRAKNPSAPPGHYYWNLFDNSTRWTAPAHFTYEQQ
                     YEQKCAFEEQRTKNYYGMDEVQAVRSGKTDVESYIKAQISKSGLSEQVEANRKEQNY
                     VDKPNNKFKLTKEKKAQWEERRQKDTERAEDEERRREAEEDEDVSDEESDEDEEE
                     FEIIGTPEREALVLREHSSTPPPEELQEOLEEPPKIPEIGLIEPPKPSIPTTGAPFG
                     AWQRVEKSEKTTEIFESPLTARFREEEEEEEKKROEPKFEFGEKTASVMTKKV
                     KGPIEFKKKTANRNIRKRE"
     gene            join(34884..34934,35157..35301,35891..36211,37573..37795,
                     AL117201.1:771..1050,AL117201.1:3076..3286,
                     AL117201.1:4137..4402,AL117201.1:4901..5020,
                     AL117201.1:5096..5173,AL117201.1:5220..5345))
```

```
CDS        /gene="F56G4.5"
           join(34884..34934,35157..35301,35891..36211,37573..37795,
           AL117201.1:771..1050,AL117201.1:3076..3286,
           AL117201.1:4137..4402,AL117201.1:4901..5020,
           AL117201.1:5096..5173,AL117201.1:5220..5345)
           /gene="F56G4.5"
           /note="similar to Thioredoxin; cDNA EST yk491h3.5 comes
           from this gene"
           /codon_start=1
           /protein_id="CAB04487.2"
           /db_xref="GI:6018392"
           /translation="MPVTEVGSLPELNNLERSDANRLIIDFFANWCGPCRMISPIF
           EQFSAEYGNATFLKYNCDVARDIVQRYNISAMPTFIFLKNRQQVDMVRGANQQAIAEK
           IRQHYSPTPANPNAASDSEKRFLEQFVKCSNVPRSYQDEVFKALARSVMPEELVGRAM
           TEGPRDEKAILKDLLHWFKTQFFTWFDRPTCPKCTLKCSTDGLQGTPTREQKEGGAS
           RVEVVICDGCNTEMRFPRYNPAKLLQTRGRCGEWANCFGLLLAALNLESRFIYDTT
           DHVWNEVYLLAEQRWCHVDPCENTMDRPLLYTRGWGKTLGYCIGYGSDHVVDVTWRYI
           WDSKKLVTQRNEVRQPVFENFLSKLNSRQAEGQTEPRKRELAVRRVCELMEMMAQEAK
           NHKIGWEKIGDDLGGRITGSEEWRREGELGESGPKLLAEPIKLAPPTGPAQNYLEFN
           YDVITDTYSQPPEIGFSAQAFELENVQRVETDWNMTYLCRKRGDAPPGNISWHFDLKS
           LKKSIEKIEIRMAGIQKFEKGKAMAIACLGDSCMRLPIDCSALTIEDPKNAEILKITA
           TLSGGEGAIGFQQAQIFRTELKRGGGARTESFSVKIWMKN"
gene       join(complement(1831..1934),complement(989..1205),
           complement(227..694),complement(283118.1:17580..17700),
           complement(283118.1:17217..17527),
           complement(283118.1:17045..17165),
           complement(283118.1:16345..16468),
           complement(283118.1:16204..16293),
           complement(283118.1:15968..16157),
           complement(283118.1:15411..15526),
           complement(283118.1:14527..14750),
           complement(283118.1:13961..14201),
           complement(283118.1:13758..13915),
           complement(283118.1:13359..13602),
           complement(283118.1:13004..13202))
           /gene="F56G4.1"
CDS        join(complement(1831..1934),complement(989..1205),
           complement(227..694),complement(283118.1:17580..17700),
           complement(283118.1:17217..17527),
           complement(283118.1:17045..17165),
           complement(283118.1:16345..16468),
           complement(283118.1:16204..16293),
           complement(283118.1:15968..16157),
           complement(283118.1:15411..15526),
           complement(283118.1:14527..14750),
           complement(283118.1:13961..14201),
           complement(283118.1:13758..13915),
           complement(283118.1:13359..13602),
           complement(283118.1:13004..13202))
           /gene="F56G4.1"
           /codon_start=1
           /protein_id="CAB04489.1"
           /db_xref="GI:3877761"
           /translation="MRVDIQCLRGLAILFVFLYHLFPLTFGNGYLGVDIFFVISGYLM
           ARNLTHMKISKISDIFRFYYRRFRRILPLYYLSVAAITIAVHVCLREFWWDVNRKYSL

alignment_scores:
   Quality:             48.00          Length:           15
   Ratio:              3.429           Gaps:              0
Percent Similarity:     93.333         Percent Identity: 60.000

alignment_block:
US-08-653-294-21 x CEF56G4      ..

Align seg 1/1  to: CEF56G4     from: 1  to: 38062

   1 AlaTyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
     ||||||:::::|||:||||||:||  ||||||| |||||||||
6216 GCCTATAATGTTTTAATACAAATAATTAAATTAAATTGAAGTAT 6260

seq_name: gb_htg1:CEY53H1_2

seq_documentation_block:
```

```
WPCOMMENT
   Sequence split into 4 fragments LOCUS CEY53H1 Accession Z99775
      Fragment Name    Begin        End
      CEY53H1_0             1      110000
      CEY53H1_1        100001      210000
      CEY53H1_2        200001      310000
      CEY53H1_3        300001      403554
   Continuation (3 of 4) of CEY53H1 from base 200001 (299775 Caenorhabditis elegans chro

alignment_scores:
   Quality:             48.00          Length:           15
   Ratio:              3.429           Gaps:              0
Percent Similarity:     93.333         Percent Identity: 60.000

alignment_block:
US-08-653-294-21 x CEY53H1_2      ..

Align seg 1/1  to: CEY53H1_2     from: 1  to: 110000

    1 AlaTyrArgLeuLeuIleLysValIleArgIleValLeuLysTyr 15
      ||||||:::::|||:||||||:||  ||||||| |||||||||
99755 GCCTATAATGTTTTAATACAAATAATTAAATTAAATTGAAGTAT 99799

seq_name: gb_pl2:AC006917

seq_documentation_block:
LOCUS       AC006917   132699 bp    DNA              PLN       07-MAY-1999
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
            I, complete sequence.
ACCESSION   AC006917
VERSION     AC006917.6  GI:4757662
KEYWORDS    HTG.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 132699)
AUTHORS     Chao,Q., Shinn,P., Dunn,P., Buehler,E., Kahn,S., Kim,C., Walker,M.,
            Williams,S., Altafi,H., Araujo,R., Conn,L., Conway,A.B.,
            Gonzalez,A., Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C.,
            Li,J., Liu,S., Luros,S., Rowley,D., Schwartz,J., Toriumi,M.,
            Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and
            Ecker,J.R.
TITLE       Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
            I
JOURNAL     Unpublished (1999)
REFERENCE   2  (bases 1 to 132699)
AUTHORS     Ecker,J.R.
TITLE       Direct Submission
JOURNAL     Submitted (25-FEB-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 132699)
AUTHORS     Ecker,J.R.
TITLE       Direct Submission
JOURNAL     Submitted (07-MAY-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT     On May 7, 1999 this sequence version replaced gi:4731042.
FEATURES             Location/Qualifiers
     source          1..132699
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
BASE COUNT   43200 a  24970 c  23491 g  41038 t
ORIGIN

alignment_scores:
   Quality:             48.00          Length:           15
```

```
              Ratio:  4.000          Gaps:        0
Percent Similarity:  80.000    Percent Identity: 66.667

alignment_block:
US-08-653-294-21 x AC006917   ..

Align seq 1/1  to: AC006917  from: 1  to: 132699

    1 AlaTyrArgLeuLeuIleLysValIleArgIleValLysLysTyr 15
      ||||||||||::::  ||||||||  |||::|||  |||:::|||
36915 GCATATCGACTGATAGCTAAGGTTATTACAATTTTGTTAGCATAT 36959

seq_name: gb_htg5:AC016012

seq_documentation_block:
LOCUS       AC016012    68676 bp    DNA       HTG      18-NOV-1999
DEFINITION  Homo sapiens clone RP11-25P5, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016012
VERSION     AC016012.1  GI:6449524
KEYWORDS    HTG; HTGS_PHASE0; NULL.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68676)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-25P5
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 68676)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Project Information
            Center project name: L4618
            Center clone name: 25_P_5
            -------------
          * NOTE: This record contains 77 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *      1    847: contig of 847 bp in length
          *                gap of unknown length
          *    848   1788: contig of 941 bp in length
          *                gap of unknown length
          *   1789   2688: contig of 900 bp in length
          *                gap of unknown length
          *   2689   3569: contig of 881 bp in length
          *                gap of unknown length
          *   3570   4470: contig of 901 bp in length
          *   4471   5369: contig of 899 bp in length
          *                gap of unknown length
          *   5370   6275: contig of 906 bp in length
          *                gap of unknown length
          *   6276   7181: contig of 906 bp in length
          *                gap of unknown length
          *   7182   8083: contig of 902 bp in length
          *                gap of unknown length
          *   8084   9000: contig of 917 bp in length
          *                gap of unknown length
          *   9001   9905: contig of 905 bp in length
          *                gap of unknown length
          *   9906  10809: contig of 904 bp in length
          *                gap of unknown length
          *  10810  11708: contig of 899 bp in length
          *                gap of unknown length
          *  11709  12585: contig of 877 bp in length
          *                gap of unknown length
          *  12586  13484: contig of 899 bp in length
          *                gap of unknown length
          *  13485  14372: contig of 888 bp in length
          *                gap of unknown length
          *  14373  15257: contig of 885 bp in length
          *                gap of unknown length
          *  15258  16170: contig of 913 bp in length
          *                gap of unknown length
          *  16171  17085: contig of 915 bp in length
          *                gap of unknown length
          *  17086  17936: contig of 851 bp in length
          *                gap of unknown length
          *  17937  18850: contig of 914 bp in length
          *                gap of unknown length
          *  18851  19734: contig of 884 bp in length
          *                gap of unknown length
          *  19735  20592: contig of 858 bp in length
          *                gap of unknown length
          *  20593  21469: contig of 877 bp in length
          *                gap of unknown length
          *  21470  22352: contig of 883 bp in length
          *                gap of unknown length
          *  22353  23231: contig of 879 bp in length
          *                gap of unknown length
          *  23232  24115: contig of 884 bp in length
          *                gap of unknown length
          *  24116  25009: contig of 894 bp in length
          *                gap of unknown length
          *  25010  25892: contig of 883 bp in length
          *                gap of unknown length
          *  25893  26788: contig of 896 bp in length
          *                gap of unknown length
          *  26789  27692: contig of 904 bp in length
          *                gap of unknown length
          *  27693  28591: contig of 899 bp in length
          *                gap of unknown length
          *  28592  29453: contig of 862 bp in length
          *                gap of unknown length
          *  29454  30353: contig of 900 bp in length
          *                gap of unknown length
          *  30354  31206: contig of 853 bp in length
          *                gap of unknown length
          *  31207  32104: contig of 898 bp in length
          *                gap of unknown length
          *  32105  33004: contig of 900 bp in length
          *                gap of unknown length
          *  33005  33920: contig of 916 bp in length
          *                gap of unknown length
          *  33921  34826: contig of 906 bp in length
          *                gap of unknown length
          *  34827  35697: contig of 871 bp in length
          *                gap of unknown length
          *  35698  36565: contig of 868 bp in length
          *                gap of unknown length
```

```
Percent Similarity:  92.857     Percent Identity:  57.143

alignment_block:
US-08-653-294-21 x AC016012/rev  ..

Align seg 1/1 to reverse of: AC016012  from: 1 to: 68676

      2 TyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
        |||||:||::||:::|||::||||||:||||::: |||||
    340 TACAGAGTATTGCTGAAGATAATGAGAATAACTTCAAATAT 299

seq_name: gb_pl1:AB017067

seq_documentation_block:
LOCUS       AB017067     83689 bp    DNA              PLN    20-NOV-1999
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MJC20,
            complete sequence.
ACCESSION   AB017067
VERSION     AB017067.1  GI:3510343
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MJC20.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (sites)
AUTHORS     Nakamura,Y.
TITLE       Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
JOURNAL     Unpublished (1998)
REFERENCE   2  (bases 1 to 83689)
AUTHORS     Nakamura,Y.
TITLE       Direct Submission
JOURNAL     Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES            Location/Qualifiers
     source         1..83689
                    /organism="Arabidopsis thaliana"
                    /strain="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="5"
                    /clone_lib="Mitsui P1"
                    /clone="MJC20"
BASE COUNT    26570 a  15544 c  15663 g  25912 t
ORIGIN

alignment_scores:
    Quality:       46.00            Length:     14
    Ratio:          3.538           Gaps:        0
    Percent Similarity:  92.857     Percent Identity:  57.143

alignment_block:
US-08-653-294-21 x AB017067  ..

Align seg 1/1 to: AB017067  from: 1 to: 83689

      2 TyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
        ::::::::||||||||||||||||||||||:::|||||||
  28816 TTCAAAATTTTAATAAAAATCATAAGTTTTGTTTTAAAATAT 28857

seq_name: gb_ba2:AF027868

seq_documentation_block:
LOCUS       AF027868     87500 bp    DNA              BCT    15-NOV-1997
DEFINITION  Bacillus subtilis chromosome region between terC and odhAB.
ACCESSION   AF027868
VERSION     AF027868.1  GI:2618993
KEYWORDS    .
```

```
*  36566   37467: contig of 902 bp in length
*                 gap of unknown length
*  37468   38354: contig of 887 bp in length
*                 gap of unknown length
*  38355   39239: contig of 885 bp in length
*                 gap of unknown length
*  39240   40149: contig of 910 bp in length
*                 gap of unknown length
*  40150   41065: contig of 916 bp in length
*                 gap of unknown length
*  41066   41943: contig of 878 bp in length
*                 gap of unknown length
*  41944   42850: contig of 907 bp in length
*                 gap of unknown length
*  42851   43745: contig of 895 bp in length
*                 gap of unknown length
*  43746   44648: contig of 903 bp in length
*                 gap of unknown length
*  44649   45522: contig of 874 bp in length
*                 gap of unknown length
*  45523   46410: contig of 888 bp in length
*                 gap of unknown length
*  46411   47280: contig of 870 bp in length
*                 gap of unknown length
*  47281   48176: contig of 896 bp in length
*                 gap of unknown length
*  48177   49093: contig of 917 bp in length
*                 gap of unknown length
*  49094   50005: contig of 912 bp in length
*                 gap of unknown length
*  50006   50869: contig of 864 bp in length
*                 gap of unknown length
*  50870   51732: contig of 863 bp in length
*                 gap of unknown length
*  51733   52632: contig of 900 bp in length
*                 gap of unknown length
*  52633   53533: contig of 901 bp in length
*                 gap of unknown length
*  53534   54410: contig of 877 bp in length
*                 gap of unknown length
*  54411   55296: contig of 886 bp in length
*                 gap of unknown length
*  55297   56193: contig of 897 bp in length
*                 gap of unknown length
*  56194   57046: contig of 853 bp in length
*                 gap of unknown length
*  57047   57932: contig of 886 bp in length
*                 gap of unknown length
*  57933   58818: contig of 886 bp in length
*                 gap of unknown length
*  58819   59728: contig of 910 bp in length
*                 gap of unknown length
*  59729   60599: contig of 871 bp in length
*                 gap of unknown length
*  60600   61503: contig of 904 bp in length
*                 gap of unknown length
*  61504   62396: contig of 893 bp in length
*                 gap of unknown length
*  62397   63302: contig of 906 bp in length
*                 gap of unknown length
*  63303   64196: contig of 894 bp in length
*                 gap of unknown length
*  64197   65059: contig of 863 bp in length
*                 gap of unknown length
*  65060   65963: contig of 904 bp in length
*                 gap of unknown length
*  65964   66855: contig of 892 bp in length
*                 gap of unknown length
*  66856   67766: contig of 911 bp in length

alignment_scores:
    Quality:   47.00        Length:    14
    Ratio:      3.615        Gaps:       0
```

```
SOURCE     Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1  (bases 59208 to 59540)
  AUTHORS   Muller,J., Walter,F., van Dijl,J.M. and Behnke,D.
  TITLE     Suppression of the growth and export defects of an Escherichia coli
            secA(Ts) mutant by a gene cloned from Bacillus subtilis
  JOURNAL   Mol. Gen. Genet. 235 (1), 89-96 (1992)
  MEDLINE   93062812
REFERENCE  2  (bases 34592 to 35233)
  AUTHORS   Wolf,M., Geczi,A., Simon,O. and Borriss,R.
  TITLE     Genes encoding xylan and beta-glucan hydrolysing enzymes in
            Bacillus subtilis: characterization, mapping and construction of
            strains deficient in lichenase, cellulase and xylanase
  JOURNAL   Microbiology 141 (Pt 2), 281-290 (1995)
  MEDLINE   9519081
REFERENCE  3  (bases 1 to 87500)
  AUTHORS   Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D.
  TITLE     Sequence analysis of the Bacillus subtilis chromosome region
            between the terC and odhAB loci cloned in a yeast artificial
            chromosome
  JOURNAL   Unpublished
REFERENCE  4  (bases 1 to 87500)
  AUTHORS   Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1997) Laboratoire de Genetique Microbienne, INRA,
            Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
FEATURES          Location/Qualifiers
     source        1..87500
                   /organism="Bacillus subtilis"
                   /db_xref="taxon:1423"
     gene          <1..560
                   /gene="yoxB"
                   /citation=[3]
     CDS           <1..560
                   /gene="yoxB"
                   /citation=[3]
                   /codon_start=3
                   /transl_table=11
                   /product="YoxB"
                   /protein_id="AAB84418.1"
                   /db_xref="GI:2618994"
                   /translation="AYETAAESLLQGGTLEQMAVDTKIAAGFSGSMRHNLITMLDLCG
                   LPQAFGIQSAAKLFGELRHMLHTTSVIKYPVFIQQKNYTGYKPAITHSPILSTYAFGH
                   FPAELNHVTGPALLIPLGKAAETVCETLLRQHSLQNLICLNGFPHPSGANGHRLKQFS
                   KNKEQLERQIRSFAALVDFAIEKRK"
     gene          604..1212
                   /gene="yoaA"
     CDS           604..1101
                   /gene="yoaA"
                   /note="similar to B.subtilis P20 protein and E.coli
                   ribosomal protein alanine acetyltransferase (194 aa)"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="putative alanine acetyl transferase"
                   /protein_id="AAB84419.1"
                   /db_xref="GI:2618995"
                   /translation="MFPILETDRLILRQITDQDAEAIFACFSNDEVIRYYGLENMESI
                   EQAISMIQTFAALYQEKRGIRWGIERRDTKELIGTIGFHALAQKHRRAEIGYEIIPEH
                   WRNGFASEVISKVVSYGFSALGLSRIGAVVFTDNEASNRLLLKMGFQKEGVLRQYMYS
                   KWDSI"
     terminator    1187..1212
                   /gene="yoaA"
     gene          complement(1217..2461)
                   /gene="yoaB"
     CDS           complement(1217..2461)
                   /gene="yoaB"
                   /note="similar to B.subtilis multidrug resistance protein
                   2 (400 aa)"
                   /citation=[3]
                   /codon_start=1
```

```
                   /transl_table=11
                   /product="putative transporter"
                   /protein_id="AAB84444.1"
                   /db_xref="GI:2619020"
                   /translation="MLDKIGIPKRLAWGFLGVVLFMGDGLEQGWLSPFLIENGLTVQ
                   QSASIFSIXGIALAIASWFSGVCLEAFGAKRTMEMGLLFYVIGTAAFIVFGFEOLNLP
                   VMVYTFVKGLGYPLFAYSFLTWVIYRTPQSKLSTAVGWFWIAYCLGMFVFGAWISSY
                   AIKAFGYLNTLWSSIFWVCLGAFFALFINKDRFEKKKRKRSETAEELLKGVTILFTNP
                   RVLTGGIIRIINSIGTYGFPVFLPMHMAQHGISTNVWLQIWGTIFLGNIVFNLIFGIV
                   GDKFGWKNTVIWFGGVGCGIFTVLLYYAPVFSGGSLAVVSVIGFIWGGLLAGYVPIGA
                   IVPTVAGKDKGAAMSVLNLAAGLSAFVGPALAWLFIGLVGAQGVWIFAALYLASAVL
                   TKCIHIPEEKAVKETSPQYAS"
     gene          complement(2555..4018)
                   /gene="yoaC"
                   /citation=[3]
     CDS           complement(2555..4018)
                   /gene="yoaC"
                   /note="similar to Lactobacillus xylulose kinase (501 aa)"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="xylulose kinase"
                   /protein_id="AAB84445.1"
                   /db_xref="GI:2619021"
                   /translation="MKKQKGYLVFDIGTGNARVAVVSVTGSVLTVEREDIEYSTETLY
                   PDSRYFSPQVLMKVQVMNLAKRALSRSCDIIGILTSTSQRGIVLIDQNGPNFLGLPN
                   IDNRGREWEAGIPDWEEIYSSTGRLPTALFSALKLYGLKRQPSLWEKTASFTSISDW
                   VTYQLSGILTYEPSQATETLLFDVKQNTWSEEMCDIFGFSPSILPPLVRAGTAIGTIA
                   NEYASELGLSINAKVIAGGDTQLAVKSTGAGLEDIVIVSGTTPITKITEDHGDTKH
                   KAWLNCHTDQGHWLVETNPGITGLNYQKLKQIFYPNETYEVMEEISALAKEDHACVA
                   ALGSYLSAEKNALTRGGFLFDAPLSAHLKRAHFVRAALEEIAFSIKWNFDILTEVTPF
                   ERDYVWCGGFQSKALTQYIADLLQKKVYVQEGYHQASVVGAAVICNETFQLTEEMS
                   ANVRVIEPKDCQIEIALYEEWKQTQRFFSGSESKVLI"
     gene          complement(4036..5070)
                   /gene="yoaD"
     CDS           complement(4036..5070)
                   /gene="yoaD"
                   /note="similar to B.subtilis D-3-phosphoglycerate
                   dehydrogenase (419 aa) encoded by serA gene"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="YoaD"
                   /protein_id="AAB84446.1"
                   /db_xref="GI:2619022"
                   /translation="MKNTMKRMFCSMTVLVTAPYNEEGRKELENLFGSVAYQSWKEQG
                   RAYREDELIQLLKATNATGLITELDQVTDSVFASVPELSVGVCRGMPSNVDVAAASX
                   RGIPVFYTPGRNAQAVAEMFIGNYISFLRHTSASNQWLKDGEWDSDYLQAYVKFKGNE
                   LTGKTVGMIGFGAVGQRIAKLLFAFDCKIKYDPYIQDDHPLYEKASLKTVFSDSDIV
                   SVHLPRTEEFLGLIDRQYFDLMKESAIFVNTSRAVVVNREDLLFVLKEHKISGAIIDV
                   FYHEPPEESDYELISLPNVLATPHLAGATFEVEDHHVTLLNKALKKWKGEKTLNIQTM
                   YNKDALKTGG"
     gene          5394..7427
                   /gene="yoaE"
     CDS           5394..7427
                   /gene="yoaE"
                   /note="similar to B.subtilis YyaE protein (667 aa) and
                   E.coli formate dehydrogenase (715 aa)"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="putative formate dehydrogenase"
                   /protein_id="AAB84420.1"
                   /db_xref="GI:2618996"
                   /translation="MRSFATQQNGIFKSVCSLDCPDQCGLLIHKKDGKIVKVQGDPDH
                   PVTAGNICNKVRNWTERIIDEKRLTTPLKRTGAKGQAIFEPISWKEAIDTITSRWKQL
                   IDEEGAESILPYSFYGNMGKLTAEGMDRRFFYRMGSSQLERTICSKAGSEGYKTMGI
                   SAGIDPEFVHTKLFIFWGINAVSTNMHQTIAQKARKGAKIVIDVHKNQTGRLAD
                   WFIPIKPGTDSALGLGIMHILFKENLHDEAFLSEYTVGIEELREHVKQYDPEKVSTIT
                   GVSTEDIYRLAKMYGETSPSFIRIGNGPHHDNGGMIVRTIACLPAITGCGWLHTGGGA
                   IKHNGILEYNTNALQRPDLLKGRTPRSFNMNQLGRVLLETDPPIRSLFIYGTNPAVV
                   APEANKVRQGLLREDLFTVVHDLFLTETAAYADIVLPATSAFENTDFYTSYWHHYIQL
                   QQPVIERYGESKSNTEVFRLLAEAMGFTDQELKDSDEVLIRQALDHPDNPHLAEIDYD
```

```
     gene          SLTKHSFMAKREKPLFPGELPTPSGKIELYSEKMKQDGFPALPTYTPLVTDNEHPFM
                   YVPGPNHNFLNSTFSNNEKHIKLEKTPKLFINTKDAEKHGIVDGAPVRINNSRGECEL
                   TAAVGEQVLPGVVVSQGLWADEQGKKTACKRANPRSLVRYGRRRHLFFRPCTD"
                   complement(9379..11027)
                   /gene="yoaH"
     terminator    complement(9379..9406)
                   /gene="yoaH"
     CDS           complement(9423..11027)
                   /gene="yoaH"
                   /note="similar to B.subtilis methyl-accepting chemotaxis
                   protein McpA (661 aa)"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="YoaH"
                   /protein_id="AAB84447.1"
                   /db_xref="GI:2619023"
                   /translation="MISSTVKKNEELKDKMEFQKEIKHIQYELTGLSNDERGFLITRD
                   KEYDEGMKGKADDVLKSLDRVNDLIDEEKYQSNIEDIKTSFTQYRALNQQVVTAYSSN
                   PKKAETIHFGEERTIRKEGVAPAVNKLSDRLDQEVEDLKDEIQGNGKMSQSLIIIVTG
                   ISVILGIVLSIMLLKSIMVPLRSINKQLEEIAHGEADLTKKVIVKNKDEFGQLAQSFN
                   SFTHSLTQIVKQISSSEOVAASSEELSASAEESKSTSEHISRAMQMAADSNVKQSSM
                   TEKSAESIITELLDSSSVASNTGNIADLSSSMRDKAEIGSKSVNKMLDQMKFIDKSVD
                   SAGNGLQTIVASTAEISDISSLITTISEQTNLLALNAAIEAARAGEQGKGFAVVAEEV
                   RKLADETNKSANHIQSVVARIQNESIETVNNIKVQENVSSGIVLSQETTGNFNEILN
                   LVEQVTSQIQEVAAATQQLISGVEVIQHTVHTLAAGTKETSANTEAVANSSQEQLHSM
                   GEISYAAESLSQLAEELQTVINRFKY"
     gene          11755..12910
                   /gene="yoaI"
     CDS           11755..12882
                   /gene="yoaI"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /note="similar to H.influenzae HI0035 protein (551 aa)"
                   /citation=[3]
                   /codon_start=1
                   /transl_table=11
                   /product="transcription antiterminator"
                   /protein_id="AAB84421.1"
                   /db_xref="GI:2618997"
                   /translation="MNTVMSFASSAELLKDKENCFPEHILDMYEQAAKHDLSFTHTF
                   ITPQVNRSQSYFGLSEKPISAKVIDRTEKGLMIHGARLLATQGGLTDEIIVFSAPKF
                   FETDEAYAFSIPSNTKGVKFITRESFVLSDSSFNHPLSSRYEEMDSIVVFDHVLVPWN
                   RVFFYDNVEAAKDFMTKSSFHAFTFHQVVIRQMIKIEFLGVAQLLVDTINVSEYQHI
                   QEKLSEIIVGLETIKALIIDKSENDAQLDEFGYMRPCLIPLQVISTIIPKLYPRFTEII
                   QLIGASGMVTLPTENAFDSEIREDLDQYLQATNTNAEERVKIFRLAWDLTMSSFGTRQ
                   THIERYFFGDPIRISSRLTITSYPKQEQLNMKTFLHADTEH"
     terminator    12885..12910
                   /gene="yoaI"
     gene          complement(12919..13617)
                   /gene="yoaI"


alignment_scores:
       Quality:    46.00         Length:     13
         Ratio:    3.538           Gaps:      0
  Percent Similarity: 100.000  Percent Identity: 69.231

alignment_block:
  US-08-653-294-21 x AF027868

  Align seg 1/1  to: AF027868  from: 1  to: 87500

      1 AlaTyrArgLeuLeuLysValIleuArgIleValLeu 13
        |||||||||||||||::|||||||||||||||||||
  35864 AGCTTTAGATTGCTCGAAAGTAATAAAAATTGTACTT 35902

seq_name: gb_pl2:ATF13C5

seq_documentation_block:
LOCUS        ATF13C5   119111 bp   DNA                PLN       23-SEP-1999
DEFINITION   Arabidopsis thaliana DNA chromosome 4, BAC clone F13C5 (ESSA
             project).
ACCESSION    AL021711
VERSION      AL021711.2  GI:5738363
KEYWORDS     .
```

```
     SOURCE        thale cress.
       ORGANISM    Arabidopsis thaliana
                   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
                   eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
                   Arabidopsis.
     REFERENCE     1  (bases 1 to 119111)
       AUTHORS     Bevan,M., Pohl,T., Weizenegger,T., Bancroft,I., Mewes,H.W.,
                   Mayer,K.F.X., Lemcke,K. and Schueller,C.
       JOURNAL     Unpublished
     REFERENCE     2  (bases 1 to 119111)
       AUTHORS     EU Arabidopsis sequencing project.
       TITLE       Direct Submission
       JOURNAL     Submitted (23-SEP-1999) MIPS, at the Max-Planck-Institut fuer
                   Biochemie. Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
                   schueller@mips.biochem.mpg.de mayer@mips.biochem.mpg.de Project
                   Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
                   Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
                   E-mail: michael.bevan@bbsrc.ac.uk
     COMMENT       On Aug 18, 1999 this sequence version replaced gi:2832611.
                   Information on performance of analysis and a more detailed
                   annotation of this entry and other sequences of chromosomes 3, 4
                   and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
     FEATURES             Location/Qualifiers
        source          1..119111
                        /organism="Arabidopsis thaliana"
                        /variety="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="4"
        misc_feature    1..4494
                        /note="positions 1-4494 are not included within BAC clone
                        F13C5"
        gene            2221..3267
                        /gene="F13C5.5"
        CDS             2221..3267
                        /gene="F13C5.5"
                        /gene="F13C5.5"
                        /note="weak similarity to MICROTUBULE ASSOCIATED PROTEIN
                        1A (FRAGMENT) - HOMO SAPIENS, EMBL:O95643"
                        /codon_start=1
                        /product="putative protein"
                        /protein_id="CAB2866.1"
                        /db_xref="GI:5738364"
                        /translation="MRWGRKKPVSSSSSSGLSRALPVSWFSKLSGSSDLKPAKEKKQD
                        EKASQNISVKTLSSTTRRSDIHENSKRFQRVSVEKNSATRSADKESNEKFEEIMSS
                        VRKKVRDFQKETCGFLEVEAMDRDNGTVILTPRIQVNRDKQRCERRDQRLEQKPKRS
                        EQDAGVKVKPARRTGTGGYSREDSVILGHTITKPAHQWEKLKEVKLREVKLKADQQR
                        KSLYLKRELNRIGTKENNKVRVFSPRASEKCRVKIDLKKAKQRAREHELLETADG
                        GMENESFAVVKCSSDPQKDFRDSMIEMIMENGINHPEELKELLVCYLRLNTDEYHDMI
                        ISVFQQVHNDFNFH"
        exon            2221..3267
                        /gene="F13C5.5"
                        /number=1
        gene            3488..5125
                        /gene="F13C5.10"
        exon            complement(3488..5125)
                        /gene="F13C5.10"
                        /number=1
        CDS             complement(3488..5125)
                        /gene="F13C5.10"
                        /note="similarity to hypothetical proteins - Arabidopsis
                        thaliana"
                        /codon_start=1
                        /product="putative protein"
                        /protein_id="CAA16741.2"
                        /db_xref="GI:5738365"
                        /db_xref="SPTREMBL:O49399"
                        /translation="MVEKDIYMCAEIIIRPQAYNLRLLQKENLKKMSVCSSTPVPILS
                        FTERAKSLTEIQQAHAFMLKTGLFHDTFSASKLVAFAATNPEPKTVSYAHSLNRIGS
                        PNGFWHNSVIRAYANSSTPEVALTVPEREMLLGPVFPDKYSPTFVLKACAAFCGFEEGR
                        QIHGLFIKSGLVTDVFVENTLVNVYGRSGYFEIARKVLDRMPVRDAVSWNSLLSAYLE
                        KGLVDEARALPDEMEERNVESWNFMISGYAAAGLVKEAKEVFDSMPVRDVVSWNAMVT
                        AYAHVGCYNEVLEVFNKMLDDSTEKPDGFTLVSVLSACASLGSLSQGEWHVYIDKHG
                        IEIEGFLATALVDMYSKCGKIDKALEVFRATSKRDVSTWNSIISDLSVHGLGKDALEI
```

```
CDS       FSEMVYEGFKPNGITFIGVLSACNHVGMLDQARKLFEMMSSVYRVEPTIEHYGCMVDL
          LGRMGKIEEAEELVNEIPADEASILLESLLGACKRFGQLEQAERIANRLLELNLRDSS
          GYAQMSNLYASDGRWEKVIDGRRNMRAERVNRSLDVA"
          complement(6832..7011)
          /gene="F13C5.20"
          /codon_start=1
          /product="hypothetical protein"
          /protein_id="CAA16742.1"
          /db_xref="GI:2832613"
          /db_xref="SPTREMBL:O49400"
          /translation="MVDLDQLQQIRDLLENIINHLEANELLPQPNPMYHDPLEDLGDS
          PASSDVTEGSNSDED"

exon      complement(6832..7011)
          /gene="F13C5.20"
          /number=1

gene      6832..7011
          /gene="F13C5.20"

gene      9456..10357
          /gene="F13C5.30"

CDS       join(9456..9560,10169..10357)
          /gene="F13C5.30"
          /codon_start=1
          /product="hypothetical protein"
          /protein_id="CAA16743.2"
          /db_xref="GI:5922639"
          /db_xref="SPTREMBL:O49401"
          /translation="MPSVQVRMDNKTAAELDAIAENLLKCCIDVMTITKHNLALQFR
          SFITNYPPKALVFDSSSAVSDLGFTESETGASTRVLGKSASWIFIFFNVTEKC"

exon      9456..9560
          /gene="F13C5.30"
          /number=1

intron    9561..10168
          /gene="F13C5.30"
          /number=1

exon      10169..10357
          /gene="F13C5.30"
          /number=2

gene      10938..11996
          /gene="F13C5.40"

CDS       join(10938..11150,11238..11639,11736..11996)
          /gene="F13C5.40"
          /note="similarity to heat shock transcription factor HSF1,
          Arabidopsis thaliana, PIR2:S52641"
          /codon_start=1
          /product="heat shock transcription factor-like protein"
          /protein_id="CAA16744.1"
          /db_xref="GI:2832615"
          /db_xref="SPTREMBL:O49402"
          /translation="MSKNEGSLTSVSYFITTIYDMVDDLSLDSIISWSQSGKSFIIWN
          PEEFYNNLLQRCFQCRINTFFSFLFSHGFRKIDSGKWEFANDNFVRQRHLINNIISD
          VIEQEVQYDQGMELFRAEKLFARLVKKVQDQLPPHNSYPTSKLPFFTKIYEMVDDPSS
          DAIISWSQSGKSFIIWNPQEFCKDHLRRLFNTLHIHFFFYKLKIFGFKKINPKKWEFA
          NDNFVRQGRHLVEIIISNDKKKNDQLRKQDAREKKMAEAGELFKLQIEEMSDMRKKMK
          KAKEVKEQEVRLCKL"

exon      10938..11150
          /gene="F13C5.40"
          /number=1

intron    11151..11237
          /gene="F13C5.40"
          /number=1

exon      11238..11639
          /gene="F13C5.40"
          /number=2

intron    11640..11735
          /gene="F13C5.40"
          /number=2

exon      11736..11996
          /gene="F13C5.40"
          /number=3

exon      complement(12538..13521)
          /gene="F13C5.50"
          /number=1

gene      12538..13820


CDS       complement(join(12538..13521,13599..13820))
          /gene="F13C5.50"
          /note="similarity to heat shock transcription factor, Zea
          mays, PIR2:S61448"
          /codon_start=1
          /product="heat shock transcription factor-like protein"
          /protein_id="CAA16745.1"
          /db_xref="GI:2832616"
          /db_xref="SPTREMBL:O49403"
          /translation="MDENNHGVSSSLPPFLTKTYEMVDDSSSDSIVSWSQSNKSFIV
          WNPPEFSRDLLPRFFKHNNFSSFIRQLNTYGFRKADPEQWEFANDDFVRGQPHLMKNI
          HRRKPVHSHSLPNLQAQLNPLIDSERVRMNNQIERLIKEKEGLLEELHKQDEREVFE
          MQVKELKERLQHMEKRQKTMVSFVSQVLEKPGLALNLSPCVPETNERKRFPRIEFFP
          DEPMLEENKTCVVVREEGSTSPSSHTREHQVEQLESSIAIWENLVSDSCESMLQSRSM
          MTLDVDESSTFPESPPLSCIQLSVDSRLKSPFSSPRIIDMNCEPDGSKEQNTVAAPPP
          PVAGANDGFWQQFSENPGSTEOREVQLERKDKDKAGVRTEKCWNSRNVNAITEQL
          GHLTSSERS"

intron    complement(13522..13598)
          /gene="F13C5.50"
          /number=1

exon      complement(13599..13820)
          /gene="F13C5.50"
          /number=2

exon      17543..17758
          /gene="F13C5.60"
          /number=1

gene      17543..20028
          /gene="F13C5.60"

CDS       join(17543..17758,19390..20028)
          /gene="F13C5.60"
          /note="similarity to predicted protein, Arabidopsis
          thaliana, PID:G3834322"
          /codon_start=1
          /product="putative protein"
          /protein_id="CAA16746.1"
          /db_xref="GI:2832617"
          /db_xref="SPTREMBL:O49404"
          /translation="MTSGTRTPTWKERENNKRRERRRAIAAKIFAGLRIHGNFKLPK
          HCDNNEVLKALCNEAGWTVEDDGTTYRKGCKPMDRMDLMNGSTSASPCSSYHGHPRAS
          YNPSPSSSSFPSPTNPFGDANSLIPWLKNLSSNSPSKLPFFHGNSISAPVTPPLARSP
          TRDQVTIPDSGWLSGMQTPQSGPSSPTFSLVSRNPFFDKEAFKMGDCNSPMWTPGQSG
          NCSPAIPAGVDQNSDVPMADGMTAEFAFGCNAMAANGMVKPWEGERIHGECVSDDLEL
          TLGNSRTR"

intron    17759..19389
          /gene="F13C5.60"
          /number=1

exon      19390..20028
          /gene="F13C5.60"
          /number=1


alignment_scores:
     Quality:       46.00          Length:      14
     Ratio:          3.286          Gaps:         0
Percent Similarity: 100.000    Percent Identity:  50.000

alignment_block:
US-08-653-294-21 x ATF13C5/rev

Align seg 1/1  to reverse of: ATF13C5  from: 1  to: 119111

       2 TyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
         ::::::::::::::|::::::::::::::::::|:::::|||
   52045 CATCATATATTAATTAAAATCATTCGTGTAATATTAAGTTAC 52004


seq_name: gb_htg6:AC011694

seq_documentation_block:
LOCUS       AC011694   160557 bp   DNA               HTG        03-DEC-1999
DEFINITION  Homo sapiens clone RP11-19D19, *** SEQUENCING IN PROGRESS ***, 33
            unordered pieces.
ACCESSION   AC011694
VERSION     AC011694.2  GI:6478993
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 160557)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens, clone RP11-19D19
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 160557)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
          Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,.
          Brown,A., Castle,A., Colangelo,A., Collins,S., Collymore,A.,
          Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
          Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
          Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
          Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
          McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
          Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,M., Talamas,J.,
          Tesfaye,S. Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,.
          Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (10-OCT-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Nov 30, 1999 this sequence version replaced gi:6018128.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ------- Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submission@genome.wi.mit.edu
          ------- Project Information
          Center project name: L3721
          Center clone name: 19_D_19
          -------
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 33 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *            1    186: contig of 186 bp in length
        *         1220:    gap of unknown length
        *         1221   2803: contig of 1034 bp in length
        *         2803:    gap of unknown length
        *         2804   3945: contig of 1583 bp in length
        *         3945:    gap of unknown length
        *         3946   5635: contig of 1142 bp in length
        *         5635:    gap of unknown length
        *         5636   6899: contig of 1690 bp in length
        *         6899:    gap of unknown length
        *         6900   8860: contig of 1264 bp in length
        *         8860:    gap of unknown length
        *         8861  10674: contig of 1961 bp in length
        *        10674:    gap of unknown length
        *        10675  13436: contig of 1814 bp in length
        *        13436:    gap of unknown length
        *        13437  16174: contig of 2762 bp in length
        *        16174:    gap of unknown length
        *        16175  19257: contig of 2738 bp in length
        *        19257:    gap of unknown length
        *        19258  22509: contig of 3083 bp in length
        *        22509:    gap of unknown length
        *        22510  25504: contig of 3252 bp in length
        *        25504:    gap of unknown length
        *        25605  28482: contig of 3095 bp in length
        *        28482:    gap of unknown length
        *        28483  31592: contig of 2878 bp in length
        *        31592:    gap of unknown length
        *        31593  34072: contig of 3110 bp in length
        *        34072:    gap of unknown length
        *        34073  38004: contig of 2480 bp in length
        *        38004:    gap of unknown length
        *        38005  43296: contig of 3932 bp in length
        *        43296:    gap of unknown length
        *        43297  48099: contig of 5292 bp in length
        *        48099:    gap of unknown length
        *        48100  52664: contig of 4803 bp in length
        *        52664:    gap of unknown length
        *        52665  56856: contig of 4565 bp in length
        *        56856:    gap of unknown length
        *        56857  61932: contig of 4192 bp in length
        *        61932:    gap of unknown length
        *        61933  67427: contig of 5076 bp in length
        *        67427:    gap of unknown length
        *        67428  71294: contig of 5495 bp in length
        *        71294:    gap of unknown length
        *        71295  76756: contig of 3867 bp in length
        *        76756:    gap of unknown length
        *        76757  82155: contig of 5462 bp in length
        *        82155:    gap of unknown length
        *        82156  90085: contig of 5399 bp in length
        *        90085:    gap of unknown length
        *        90086  98690: contig of 7930 bp in length
        *        98690:    gap of unknown length
        *        98691 107744: contig of 8605 bp in length
        *       107744:    gap of unknown length
        *       107745 116715: contig of 9054 bp in length
        *       116715:    gap of unknown length
        *       116716 132246: contig of 8971 bp in length
        *       132246:    gap of unknown length
        *       132247 145974: contig of 15531 bp in length
        *       145974:    gap of unknown length
        *       145975 160557: contig of 13728 bp in length
        *                   gap of unknown length
        *                   contig of 14583 bp in length.

FEATURES          Location/Qualifiers
     source        1..160557
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="RP11-19D19"
                   /clone_lib="RPCI-11 Human Male BAC"

BASE COUNT    44509 a  35116 c  36309 g  44566 t    57 others
ORIGIN

alignment_scores:
        Quality:       46.00           Length:    15
        Ratio:          3.538          Gaps:       0
Percent Similarity:  86.667  Percent Identity: 60.000

alignment_block:
US-08-653-294-21 x AC011694/rev

Align seg 1/1  to reverse of: AC011694  from:  1   to: 160557

      1 AlaTyrArgLeuLeuLysValIleArgIleValLeuLysTyr 15
        |||||||||::::|||||          :::||:::|||::::|||||||
  88764 GCATATAGATTCTAATATGTTTGTTAGACTTACGCTCAAGTAT 88720

seq_name: gb_bal:BSUB0011

seq_documentation_block:
LOCUS       BSUB0011    207730 bp    DNA              BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 11 of 21): from 2000171
            to 2207900.
ACCESSION   Z99114 AL009126
VERSION     Z99114.1  GI:2634230
KEYWORDS    .
SOURCE      Bacillus subtilis.
ORGANISM    Bacillus subtilis
```

```
REFERENCE              Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
  AUTHORS              Bacillus.
                       1 (bases 1 to 207730)
                       Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
                       Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
                       Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
                       Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
                       Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
                       Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
                       Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
                       Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
                       Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
                       Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
                       Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
                       Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
                       Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
                       Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
                       Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
                       Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
                       Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
                       Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
                       Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
                       Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
                       Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
                       Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
                       Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
                       Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
                       Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
                       Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
                       Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
                       Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
                       Yoshida,H., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
                       Danchin,A.
  TITLE                The complete genome sequence of the gram-positive bacterium
                       Bacillus subtilis
  JOURNAL              Nature 390 (6657), 249-256 (1997)
  MEDLINE              98044033
REFERENCE              2 (bases 1 to 207730)
  AUTHORS              Kunst,F. Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE                Direct Submission
  JOURNAL              Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
                       Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
                       Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
                       adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
                       68 89 48
FEATURES               Location/Qualifiers
     source            1..207730
                       /organism="Bacillus subtilis"
                       /strain="168"
                       /db_xref="taxon:1423"
     gene              complement(10..1436)
                       /gene="yoeA"
     terminator        complement(10..33)
                       /gene="yoeA"
     CDS               complement(45..1436)
                       /gene="yoeA"
                       /function="unknown"
                       /note="similar to hypothetical proteins from B. subtilis" .
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAB13730.1"
                       /db_xref="GI:2634231"
                       /db_xref="SPTREMBL:O34474"
                       /translation="METSQEPLVTQKKSKPSVWRSMSLFLVPLLLSNVLQSVGQLVGM
                       MAVGRWLGVDAVAAVSSFFPLFLLSFTIGISGGSILLIGQAYGAKNEERLKAVVGT
                       KTKDTTIIDSRERAPAGATPDMFLDENGKAIPFSERVTKGTAVGVPGTLKGLEEALDK
                       WGTRSMQLITPSIKLAEGFPIDSVLAEAISDYQEKLSRTAAKDVFLPNGEPLKEGD
                       TLIQKDLAKTFKLIRSKGTDAFYKGFKAFTLSDTVQDFGGSMTEKDLENYDITIDEPI
                       WGDYQGYQIATPFPPSSGGIFLQMKILDHFNLSQYDVRSWEKYQLLAETMHLSYAD
                       RASYAGDPEFVNPLKGLLHPDYIKERQQLINDQVNKKPKAGDPWKYQEGSANYKQV
                       TFLRGTGDSKIPFYTLLIVSTVINIALLPVLLGMFGFPKLGIYGSAYATVISTIATFL
                       VLMVYLRRKHPLQFDKTVRRYLKMDKELVLLLRLGVPASINMIUVSLSEIAVISFV
                       NHYGSNATAAYGVVNQVASIYQMPAVSLGIAVSIFAAQSIGANEFDRLKQVIRVGIWL
                       NYIIGGVLIILYVFSHQILSLFLTEQESLYIAHRLLMTLIWSYLLFGNAQIISATMR
                       AAVEEPRIYTNSMSSYRYEDGVPKDVLSKLNGMGHKFGTSPVDIGNVQSISIDHENGT
                       FKGVADSSRNGAAIGINLKRK"
                       FVWKKKQITRLIQ"
     gene              1722..2345
```

```
  CDS               1722..2267
                    /gene="yoeB"
                    /gene="yoeB"
                    /function="unknown"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB13731.1"
                    /db_xref="GI:2634232"
                    /db_xref="SPTREMBL:O34841"
                    /translation="MKKCLLFLTTIALILSLSTNAFAKNTSGDLSQKQALQALSARE
                    HFWNTMSGQNPKVKAVCPSGTFEYQNLQYYMCSDLGTKAKAVNYLTPIFTKTAIEK
                    GFKDYHFTVSKGKLAVPIGDGDNLLNWKKSTAKLISKKGSTITYEFTVPTLDGSPSAK
                    RKVTFVKENKKWKVNQFDAVI"
     terminator     2274..2292
                    /gene="yoeB"
     terminator     2319..2345
                    /gene="yoeB"
     tRNA           2359..2434
                    /gene="trnSL-Arg1"
                    /product="transfer RNA-Arg"
     gene           2359..2434
                    /gene="trnSL-Arg1"
     gene           complement(2745..3032)
                    /gene="yoeC"
     CDS            complement(2745..3032)
                    /gene="yoeC"
                    /function="unknown"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB13732.1"
                    /db_xref="GI:2634233"
                    /db_xref="SPTREMBL:O35009"
                    /translation="MHIVQPIRSLEKIQEVKQYLLNKNKRDYFLFIGINSALRISDI
                    LPLQVRDVQNKDHLWATESQNEKEKKDSYFRIAETGNIRVYERHEGKRISF"
     terminator     complement(3327..3347)
                    /gene="yoeD"
     gene           complement(3327..3578)
                    /gene="yoeD"
     gene           complement(3348..3578)
                    /gene="yoeD"
     CDS            complement(3348..3578)
                    /gene="yoeD"
                    /function="unknown"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAB13733.1"
                    /db_xref="GI:2634234"
                    /db_xref="SPTREMBL:O34555"
                    /translation="MYLTIEETAEYTNLSEDYIKSLVLEGKIRAVHDGEQFLIYKEQF
                    KTHLEQLENYKALVQEILNEFIPEDIDVKDED"
     gene           3763..5526
                    /gene="ggt"
     gene           3763..5526
                    /gene="ggt"
                    /EC_number="2.3.2.2"
                    /function="glutathione metabolism"
                    /note="alternate gene name: pac"
                    /codon_start=1
                    /transl_table=11
                    /product="gamma-glutamyltranspeptidase"
                    /protein_id="CAB13734.1"
                    /db_xref="GI:2634235"
                    /db_xref="SWISS-PROT:P54422"
                    /translation="MKRTWNVCLTALLSVLLVAGSVPFHAEAKKPPKSYDEYKQVDVG
                    KDGMVATAHPLASEIGADVLRKGGNAIDAAVAIQFALNVTEPMSGIGGGGFMMVYDG
                    ...
```

     gene

```
terminator        /gene="yofA"
                  complement(5535. .5559)
CDS               /gene="yofA"
                  complement(5627. .6484)
                  /gene="yofA"
                  /function="unknown"
                  /note="similar to transcriptional regulator (LysR family)"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAB13735.1"
                  /db_xref="GI:2634236"
                  /db_xref="SPTREMBL:O34685"
                  /translation="MESGDLKIFQAVAREGSITKAAQMLNYVQSNVTARVHNLEEDLN
                  IRLFHRTNRGMKLTAAGENLLQYADQVLSLLDQAEKSTRMSRQPKGPLRIGSLETMAV
                  THLPEHAASFLRRFPEVDLSVNTADTHHLIQQVLDHKVDGAFVYGPVEHAAVRQLHVS
                  HDELVLISSREGTAEDMLQQPMLFFGAGCSHRDRVKRLLEEAGIHNQKIIEFGTLEAI
                  IKGVSAGMGTALLPKSAVDGSEHRTNVWIHQLPPSYQDLEIVFIYRKDFFITSAFQTF
                  LDEINEMKR"
gene              6612. .7601
                  /gene="yogA"
CDS               6612. .7601
                  /gene="yogA"
                  /function="unknown"
                  /note="similar to alcohol dehydrogenase"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAB13736.1"
                  /db_xref="GI:2634237"
                  /db_xref="SPTREMBL:O35017"
                  /translation="MKAVIHNGKAGLIGLSVODVPSTKPGYGEVKVKLKSAGLNHRDL
                  FLMKNKSEQDPRMILGSDGAGIIEEIGEGVKNVTVQTEVVIFPTLNWDLTENVPPVPE
                  ILGGPSDGTLAEVVIIPSQNAIKKPAYLSWEEAGVLPLSALTAYRALFTKGQLKKGEH
                  LLIPIGSGVATYALFMAKAIGATVSVTSRSEEKRKKALKLGADYAFDSYSNWDEQLQ
                  GKKIDVVLDSIGPALFSEYFRHVKPNGRIVSFGASSGDNLSFPVRSLFFPQVNVLGTS
                  MGSGEEFQAMLAFIDKHLRPVIDRIIYPLEKACEAYKRMQEGRQFGNIGIVME"
gene              complement(7611. .9139)
                  /gene="gltB"
terminator        complement(7611. .7625)
                  /gene="gltB"
CDS               complement(7658. .9139)
                  /gene="gltB"
                  /EC_number="1.4.1.13"
                  /function="glutamate biosynthesis"
                  /codon_start=1
                  /transl_table=11
                  /product="glutamate synthase (small subunit)"
```

```
alignment_scores:
        Quality:      46.00        Length:      13
        Ratio:         3.538        Gaps:         0
  Percent Similarity: 100.000    Percent Identity: 69.231

alignment_block:
US-08-653-294-21 x BSUB0011

Align seq 1/1  to: BSUB0011  from: 1  to: 207730

   1 AlaTyrArgLeuLeuLysValIleArgIleValLeu 13
     :::::|||||||:|||::|||||:|||||:|||||||
54955 AGCTTTAGATTGCTGCTGAAAGTAATAAAAATTGTACTT 54993
```

```
REFERENCE   1  (bases 1 to 8642)
  AUTHORS   Bertrand,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1990) Bertrand H., Department of Microbiology,
            University of Guelph, Guelph Ontario N1G 2W1, Canada
REFERENCE   2  (bases 1 to 8642)
  AUTHORS   Court,D.A. and Bertrand,H.
  TITLE     A method for sequencing uncloned termini of linear plasmids
  JOURNAL   Nucleic Acids Res. 19 (7), 1714 (1991)
  MEDLINE   91227177
REFERENCE   3  (bases 1 to 8642)
  AUTHORS   Chan,B.S., Court,D.A., Vierula,P.J. and Bertrand,H.
  TITLE     The kalilo linear senescence-inducing plasmid of Neurospora is an
            invertron and encodes DNA and RNA polymerases
  JOURNAL   Curr.Genet. 20 (3), 225-237 (1991)
  MEDLINE   92035090
REFERENCE   4  (bases 7161 to 7293)
  AUTHORS   Vierula,P.J. and Bertrand,H.
  TITLE     A deletion derivative of the kalilo senescence plasmid forms
            hairpin and duplex DNA structures in the mitochondria of Neurospora
  JOURNAL   Mol. Gen. Genet. 234 (3), 361-368 (1992)
  MEDLINE   93024309
FEATURES             Location/Qualifiers
     source          1. .8642
                     /organism="Neurospora intermedia"
                     /mitochondrion
                     /db_xref="taxon:5142"
                     /plasmid="Mitochondrial plasmid"
                     /tissue_type="mycelium"
                     /cell_line="P551"
     repeat_unit     1. .1365
                     /note="inverted repeat A"
     misc_feature    1
                     /note="site of terminal protein"
     CDS             1541. .4222
                     /codon_start=1
                     /transl_table=4
                     /product="DNA polymerase"
                     /protein_id="CAA36327.1"
                     /db_xref="GI:578157"
                     /db_xref="SWISS-PROT:P33538"
                     /translation="MMNLRGFISLYKKQFKEEKFFMAKIKFNNNDIRSISTVQIGST
                     DPLEVLRLLEAITSIYMHTHGISAVSEPFEYELFSLGDGLPKGNIIFTFKPTSNPS
                     IKTKYEAHKSNIKRNKNINLSKKNPLNKFKNYGVTIPNTMDLSQWPNIHFINDGKNAV
                     SLNNIIKSGVDNMTLSFFITINKKYNEITVLLNNTPIFKIKDEKIMSEDDLSSFKKTI
                     TENEQDKVYVFENGEMVFFSENVKTSFIIKITRQDLINFENPKIITLDLETRSVPIHP
                     IKEGKDGKEPGKVDSIMFPLMSVYNGKFVKSFLFSQSAWETEMNNAFKSIMLRKYDGY
                     KVYTHNFSYFDGIFIIDILSRLGEVKPFMRNGKILKLTFNFTLPNSKRKYTLYFMDSL
                     LILPDSLDKLSNFFNNKVKKLFFPHSFLDDNTIPINYVGKCPDYKFYFPKAYTEDFTIE
                     QYQEYANKFKNNNWDLKKELIKYCEIDTIALYQVLVSFQRKIYEKFMIDCTKYPTIPS
                     LAFAIFRKKYLVEDMIPNIKSKLHNIIKLSYFGGICELYKPFGVNIKSYDVNSLYPFA
                     MKYFKMPSGIPKYVKGTLQNIVRFTDSICVPFGFYNVKVKTPLNLDKFPLPTRLNTP
                     AGRRTAPFLGQWEGWYFSEEILNAMKHGYEFEFIEGYLFEESSMFDEYIDLLYNIKKN
                     SPKESPWYYISKLLMNSLYGRFGLNSEGEEIFITSEEEGDAIIATKEYVITPLSSGN
                     VLISAKLPEAFGDMNISVPISSAAAYSRIHMSHFLTKYSNNIYYIDTGIKVDIDL
                     DKDEVDSKELGKMKYEYVFEEYTSLGPKVYGGLLYDKKGKLIELVKLRGYSSKLPYNK
     CDS             complement(4899. .7334)
                     /codon_start=1
                     /transl_table=4
                     /product="RNA polymerase"
                     /protein_id="CAA36326.1"
                     /db_xref="GI:578158"
                     /db_xref="SWISS-PROT:P33541"
                     /translation="MSKFSFFNYRIMRNSHATSLRFIKFNYSEFKNIRNFSSEKKRD
                     PKLSFLNNVSKILSEHKDLHKAQEIIEREWLKATPSDVKSDITSTMYTKHRLIH
                     DRAFDLLYKSNGNLNKISKSLDYIVNEKFIFAWSYLLASYSKSSLTNLDYRIGY
                     SITRHIYTEYKDEFMSFDDFSNYNKFDVGFFVKLGDIFINTTSPLNPIFERVFENE
                     IYSLKINKEYENEIMESLVISPKALPMVCPPLPWGHGIRGGNLINLTDQDESLVVGS
                     FHHRHKIAITDKLYNVINKLNAFFKINGDLLSVLYQNEGSFILDFYKNTKKDTYINNM
                     ITIDIARTYLNTPFYLNVNIDWRGRIYTQSFYLDYQGSELSLALINLFEGKKLDEKGL
                     FFFVYGANIYNDGGKFSKKSFQDRFNWVENLDNIIAMDKEFILKAESPTLFAAFCL
```

```
                TMRKLKENPDYPVENPIFLDATCSGVQHFAAMLLDELGKYVNLINSGESVNDFYSQL
                IPAINKAINESAEKKFKNLKFSDISLNRSLLKKVIMTKSYNVTTYGITEQLKSKLEKV
                EKIVISKGKEIKVDYLVPTKNGDFVLDTFEVETLASIINDNIFNQFPKLHSIYDYL
                TRLAKIYLKLDIPLSWSTPDGLELTQRYNLSKVKKLTINFLGKNRTAVLRSWVNEKDS
                RREVQAIIPNIIHSLDASHLTMIDSWDSYILPIHECFGTHPNDMYKLAEQVRECFIL
                LYSKNDFLNKIDYKFRENLKDYKIEIVNKNGEDFVKIKGNKKRYEYLPLPVLPQMGELN
                VEDIRDMGKYMIS"
repeat_unit     7278..8642
                /note="inverted repeat A'"
misc_feature    8642
                /note="site of terminal protein"
BASE COUNT     3042 a   1327 c   1277 g   2996 t
ORIGIN

alignment_scores:
    Quality:       45.00          Length:        12
    Ratio:          3.750          Gaps:           0
    Percent Similarity: 100.000    Percent Identity: 75.000

alignment_block:
US-08-653-294-21 x NIMIDRPM/rev

Align seg 1/1  to reverse of: NIMIDRPM  from: 1  to: 8642

     3 ArgLeuLeuLysValIleArgIleValLeuLeuLys 14
       ||||||||||||||||::|||::|||::||||||
  5041 AGATTGTTAATAAAATGGTGAGGATTTTGTTAAAA 5006

seq_name: gb_pl1:SC9571x

seq_documentation_block:
LOCUS       SC9571x        29366 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome XIII cosmid 9571.
ACCESSION   Z49810 271257
VERSION     Z49810.1  GI:854810
KEYWORDS    ERG6; lactoyl glutathione lyase; MRPL39; PAR1; PDR4; PPz1; protein
            phosphatase; ribosomal protein; SED6; SNQ3; SPT5; transcription
            initiation protein; transfer RNA-Ser; YAP1.
SOURCE      baker's yeast.
ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 29366)
AUTHORS     Gentles,S. and Bowman,S.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 29366)
AUTHORS     Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1995) Saccharomyces cerevisiae chromosome XIII
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            All CDS over 100 codons have been analysed. CDS that are completely
            overlapped and those that are overlapped by more than 50%
            of their length by a larger CDS have been omitted from this
            analysis.
            Details of the omitted CDS are available on request. The more
            significant matches with motifs in the PROSITE database are
            also included but some of these may be fortuitous. The length in
            codons and the calculated codon adaptation index (CAI)
            is given for each CDS.
            Cosmid 9571 overlapped at 5' end by cosmid 8337, EMBL entry SC8337,
            accesion no. 246659, and at the 3' end by cosmid 8270, EMBL entry
            SC8270, accession no.Z48613.
FEATURES             Location/Qualifiers
     source          1..29366
                     /organism="Saccharomyces cerevisiae"
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XIII"
                     /clone="cosmid 9571"
                     /map="13L"
```

```
misc_feature    1..463
                /note="overlap with cosmid 8337, positions 34044 34506
                EMBL:SC8337, 246659"
CDS             1057..2793
                /note="YM9571.01, unknown, len: 578, CAI: 0.14"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89935.1"
                /db_xref="GI:854473"
                /db_xref="SWISS-PROT:P50109"
                /translation="MSLEEFLGNDTLGESVWDEEDINLDAISNTTNIDILKQTKAGEH
                QRDGHQQHPHGGHGPMNRSFSNAGPFGGSMGDFANHHHPLQHQQGPPYIVKFSDLP
                PRFSNFDIEDLFQAKFTKFIKFKLFWEINKNPSISTLKSGSIFDQNFKRDSKVAFVEL
                YTSRDMDKILNYWTTPLKEIYHITTAPAEFDFKDYSTKVLLTDPKDDAGKPFITKT
                QRSKSNPFGSAKPVDTQSKILDIEEKMENLHVEDTTLRASLIPSSDSMATTATGSKI
                TILKQTPTEEESHSATPTPKPLSYSEVVERKSVVNETSKKGTPLSKLKDSPALELQSKP
                DKSDEFKGGDDEQGFEKGGDDKAQLDVSNDKGSETDVDKQFTFKNVEREHSMSRTKY
                NGNHNNNGNFRGSNRYRGGPNGSSYKGGHNNRGNGGYRGGSYNNNNNTNDNNNN
                NNNSSSNNNNGSRYHDRQNNEEGLTSDSSLDASGNKKNDFTNSTSNTQQYSIFKPASG
                FLGQGNNDSIRNNGRGNYNSSGMNGSRGRGFGRGRGFGRGAYNNRGSRGRGSSGNY
                SNYNNRTTDMPL"
gene            complement(3520..5598)
                /gene="PPz1"
CDS             complement(3520..5598)
                /gene="PPz1"
                /note="YM9571.02c, PPz1 gene, len: 692, CAI: 0.17,
                SW:PPz1_YEAST P26570 serine/threonine protein phosphatase
                PP-Z1; contains PS00125 Serine/threonine specific protein
                phosphatases signature"
                /codon_start=1
                /db_xref="SGD:S0004478"
                /product="Ppz1p"
                /protein_id="CAA89936.1"
                /db_xref="GI:854474"
                /db_xref="SWISS-PROT:P26570"
                /translation="MGNSSKSKSKRDSHSNSSSRNRPQVSRTESHSVKSAKSNRKSS
                RSRRSLPSSSTNTNSNVPDDFSPSETKAHSSHTNRYHFPSSSHSHSNSQNE
                LLTTPSSSSTKRPSTSRRSSYNTKAAADLPPSMIQMEPKSPILKTNNSSTHVSKHKSS
                YSSTYYENALIDDDNDDKDNDISHTKRFSRSSNSRSSIRSGSVSRRKSDVTHEEPNN
                GSYSSNNQENYLVQALTRSNSHASSLHSRKSSFGSDGNTAYSTPLNSPGLSKLTDHSG
                EYFTSNSTSSLNHHSSRDIYPSKHISNDDDIENSSQLSNIHASMENVNDKNNNITDSK
                KDPNEEFNDIMQSSGNKNAPKFKKPIDIDETIQKLLDAGYAAKRTKNVCLKNNEILQ
                ICIKAREIFLSQPSLELSPPVKIVSDVHGYGDLLRLFTKCGFPPSSNYLFLGDYVD
                RGKQSLETILLFCYKIKYPENFFLRGNHECANVTRVYGFYDECKRRCNIKWKTFI
                DTFNILPLAAIVAGKIFCVHGGLSPVLNSMDEIRHVVRPTDVPDFGLINDLLWSDPTD
                SPNEWEDNERGVSYCVNKVAINKFLNKFGFDLVCRAHMVVEDGYEFFNDRSLVTVFSA
                PNYCGEFDNWGAVMSVSEGLLCSFELLDPLDSAALKQVMKKGRQERKLANQQQOMET
                SITNDNESQQ"
misc_feature    complement(4156..4173)
                /gene="PPz1"
                /note="PS00125 Serine/threonine specific protein
                phosphatases signature"
                /db_xref="SGD:S0004478"
CDS             complement(6051..7091)
                /note="YM9571.03c, unknown, len: 346, CAI: 0.16"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89937.1"
                /db_xref="GI:854475"
                /db_xref="SWISS-PROT:Q04226"
                /translation="MTEPQGPIDTIPKVNYPPILTIANYFSTKQMIDQVISEDQDYVT
                WKLQNLRTGGTSINNQLNKYPKYKYQKTRINQQDPSINKVPENLIFPQDILQQQTQN
                SNYEDTNTNEDDENEKLAQDEQFKLLVTNLDKDQTNRFEVFIRTSLNKTQVKKLASTVA
                NQTISENIRVFLQAVGKIYAGBIIELAMIVVNKWLTSQMCIEFDKRTKIGVKLKKYLK
                KLTFSIIENQYKQDYQSDSVPEDEPDFYFDDEEVDKRETTLGNSLLQSKSLQQSDHN
                SQDLKLQLIEQYNKLVLQFNKLDVSIEKYNNSPLLPEHIREAWRLYRLQSDTLPNAYW
                RTQGEGQGSMFR"
                7287..8126
                /note="YM9571.04, unknown, len: 279, CAI: 0.13, similar to
                WP:C14B1.5 CE00902 Diphtheria toxin resistance protein
                (39.4% identity in 155 aa overlap) and WP:C35D10.12
                CE01193, (30.8% identity in 208 aa overlap)"
```

```
                 /codon_start=1
                 /product="unknown"
                 /protein_id="CAA89938.1"
                 /db_xref="GI:854476"
                 /db_xref="SWISS-PROT:P49957"
                 /translation="MEINQAAEKEQEYVHKVYNEIAPHFSQTRYKPWPIVTQFLKTRP
                 MGSIGIDVGCGNGKYLGVNPDIYIIGSDRSDGLIECARGINPSYNLLVADGLNLPHKN
                 ETFDFAISIAVVHHWSRERRVEVIRHVLSKLRQGGQALIYCWALEQGSSRRGYHEGM
                 EQDVFVPWVLPKSKSKPKTKSTPPAKVTRPKPNLMNIPPKERSEYLQRWKEEQQRSK
                 SLDDNDEKQQQDQEQEREEVKRYRYYHLYREGELAEDCRQAGAAVHSEGFERDNWVVA
                 QKR"
CDS              8211..9965
                 /note="YM9571.05, unknown, len: 584, CAI: 0.14, contains
                 PS00225 Crystallins beta and gamma 'Greek key' motif
                 signature"
                 /codon_start=1
                 /product="unknown"
                 /protein_id="CAA89939.1"
                 /db_xref="GI:854477"
                 /db_xref="SWISS-PROT:004228"
                 /translation="MPVVNHEDSEFHLSHTEDKLNEFQVITNFPPEDLPDVVRLLRN
                 HGWQLEPALSRYFDGEWKGEPDQMGEPTQSTPMAETLVPPALGPRPLLFTASLPVVR
                 PLPANFRNDFRTIGLNGRSNTVWSMFESFSYDGNPFFLFLILLIPRIINRLSATIFTFF
                 CTLLSLHSISGGGNSGKPKISKVPKAPTRETHIPLAEILGDTKDKDAFCELKSFKPDI
                 SFNEALRIAKEFKFMLLILVGDTPDIFTDVDVNSKLLLEKILLNKKTLQYLRKIDN
                 DLIIYLKCVHELEEPWLVARQLGVRNTPEIFLIANVANKASHSETLPSQRLSIGKLKV
                 NSLNRFLQSLTNVVEKYTPELVVNKTEMHELRMSREIKKLQEDAYKKSLEMDRIKAIE
                 KEKSLKHAQDLKLNSTARQLKWLKACIDEIQPFETTGKQATLQFRTSSGKRFVKKFPS
                 MTTLYQYQSIGCHIYLAVYSSDPAEWSNALQDKIRQLSADDMLCFKEGQLETATAT
                 TIEELGHIINNELTSFDLERGKLEFDFELVSFPFPKYTVHPNEHMSVDQVPQLWPNGSL
                 LVEALDEDEDEENEEQ"
CDS              10178..10813
                 /note="YM9571.06, len: 211, CAI: 0.22, potential
                 transmembrane protein, similar at C-terminus to
                 SW:G25L_CANFA P27869 glycoprotein 25L precursor (24.4%
                 identity in 168 aa overlap) and to SW:P24B_YEAST P32803
                 P24B protein precursor (26.0% identity in 131 aa overlap)"
                 /codon_start=1
                 /product="unknown"
                 /protein_id="CAA89940.1"
                 /db_xref="GI:854478"
                 /db_xref="SWISS-PROT:P54837"
                 /translation="MOVLQLWLTTLISLVVAVQGLHFDIAASTDPEQVCIRDFVTEGQ
                 LVVADIHSDGSVGDGQKLNLFVRDSVGNEYRKRDFAGDVVRAFTAPSSTAFDVCFEN
                 QAQYRGRSLSRAIELDIESGAEARDWNKISANEKLKPIEVELRRVEEITDEIVDELTY
                 LKNREERLRDTNESTNRRVRNFSILVIVLSSLGVWQVNYLKNYFKTKHII"
CDS              complement(10957..11490)
                 /note="YM9571.07c, unknown, len: 177, CAI: 0.12"
                 /codon_start=1
                 /product="unknown"
                 /protein_id="CAA89941.1"
                 /db_xref="GI:854479"
                 /db_xref="SWISS-PROT:004231"
                 /translation="MSKSTNVSYERVELFENPKVPIEVEDEILEKYAESSLDHDMTVN
                 ELPRFFKDLQLEPTIWKLVRNEDVIIEGTDVIDFTKIVRCTCQLLILMNNLTVIDDLW
                 SMLIRNCGRDVDFPQVALRDHVLSVKDLQKISNLIGADQSSGTIEMISCATDGKRLFM
                 TYLDFGCVLGKLGYLKM"
gene             11739..14930
                 /gene="SPT5"
                 /db_xref="SGD:S000470"
CDS              11739..14930
                 /gene="SPT5"
                 /note="YM9571.08, SPT5 gene, len: 1063, CAI: 0.21,
                 SW:SPT5_YEAST P27692 transcription initiation protein"
                 /codon_start=1
                 /product="Spt5p"
                 /protein_id="CAA89942.1"
                 /db_xref="GI:854480"
                 /db_xref="SWISS-PROT:P27692"
                 /translation="MSDNSDFNVSMQDHDQQFADPVVVPQSTDTKDENTSDKDTVDSG
                 NVTTTESTERAESTSNIPPLDGEEKEAKSEPQQPEDNAETARATEQVSSSNGPATDDAQ
                 ATLNTDSSEANEIVKKEEGSDERKRPREEDTKNSDGDTKDEGDNKDEDDDEDDDDDDD
```

```
                 DEDDDDEAPTKRRRQERNRFLDIEAEVSDDEDEDEDEDSELVREGFITHGDDEDDEA
                 SAPGARRDDRLHRQLDQDLNKTSEEDAQRLAKELRERYGRSSSKQYRAAAQDGYVPQR
                 FLLPSVDTATIWGVRCRPGKEKELIRKLLKKKFNLDRAMGKKLKILSIFQRDNYTGR
                 IYIEAPKQSVIEKFCNGVPDIYISQKLLIPVQELPLLLKPNKSDDVALEEGSYVRIKR
                 GIYKGDLAMVDQISENNLEVMLKIVPRLDYGKFDEIDPTQQRKSRRPTFAHRAPPQL
                 FNPTMALRLDQANLYKRDDRHFTYKNEDYIDGYLYKSFRIQHVETKNIQPTVEELARF
                 GSKEGAVDLTSVSQSIKKAQAAKVTFQPGDRIEVLNGEQRGSKGIVTRTKDIATIKL
                 ITANNLSKSIDTTATSSEYALHDIVELSAKNVACIQAGHDIFKVIDETGKVSTITKG
                 SILSKINTARARVSSVDANGNEIKIGDTIVEKVGSRREGQVLYIQTQQIFVSKKIVE
                 NAGVFVVNPSNVEAVASKDNMLSNKMDLSKMNPEIISKMGPPSSKTFQQPIQSRGGRE
                 VALGKTVRIRSAGYKGQLGIVKDVNGDKRATVELHSKNKHITIDKHKLTYNREGGEGI
                 TYDELVNRRGRVPQARMGPSYVSAPRNMATGGIAAGAATSSGLSGGMTPGWSSFDGG

alignment_scores:
        Quality:       45.00            Length:          14
        Ratio:          3.462           Gaps:             0
    Percent Similarity:  92.857     Percent Identity:  57.143

alignment_block:
US-08-653-294-21 x SC9571X

Align seg 1/1  to: SC9571X  from: 1  to: 29366

    2 TyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
      |||||||||||||  |||||||  |||||||||||||||||
23408 TATCGGCTACTCGTTCAAAAAATTAAAATCGTTATCAATTAT 23449

seq_name: gb_htg5:AC015353

seq_documentation_block:
LOCUS       AC015353   29459 bp    DNA        HTG        16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015353
VERSION     AC015353.1  GI:6435982
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 29459)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210956 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES           Location/Qualifiers
     source        1..29459
                   /organism="Drosophila melanogaster"
                   /db_xref="taxon:7227"
BASE COUNT  7846 a  6592 c  6879 g  8142 t
ORIGIN

alignment_scores:
        Quality:       45.00            Length:          12
        Ratio:          3.750           Gaps:             0
    Percent Similarity: 100.000     Percent Identity:  58.333

alignment_block:
US-08-653-294-21 x AC015353/rev

Align seg 1/1  to reverse of: AC015353  from: 1  to: 29459

    1 AlaTyrArgLeuIleLysValIleArgIleVal 12
      ||||||||||||||||||||||||||||
```

1853 GCCTACGGCCTCGTCGTACGGATCGTTCGGATCGTC 1818

seq_name: gb_pl1:SPAC9G1

```
seq_documentation_block:
LOCUS       SPAC9G1     30985 bp    DNA              PLN       09-SEP-1998
DEFINITION  S.pombe chromosome I cosmid c9G1.
ACCESSION   298763
VERSION     298763.1  GI:2370549
KEYWORDS    beta-transducin; cdc12 homolog; cytochrome oxidase biogenesis
            protein: Homol D box; inositol metabolism; Inositol polyphosphate
            phosphatase; lysophospholipase; MAP Kinase kinase kinase; oxa1;
            ribosomal protein L30e; rpl30; septin homolog; serine threonine
            protein kinase; spn4; src homology domain; vacuolar sorting; wak1;
            WD repeat.
SOURCE      fission yeast.
ORGANISM    Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Archiascomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 30985)
AUTHORS     Churcher,C.M. and Gentles,S.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 30985)
AUTHORS     Barrel,B.G., Rajandream,M.A. and Wood,V.
TITLE       Direct Submission
JOURNAL     Submitted (26-AUG-1997) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Spjsplice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites. CDS are numbered using the following
            system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10
            (cosmid name),.01 (first CDS), c (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid c9G1
            overlaps cosmid c3H1 at its 3' end.
FEATURES             Location/Qualifiers
     source          1..30985
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /clone="cosmid c9G1"
                     /map="I"
     gene            1..476
                     /gene="SPAC9G1.01"
     misc_feature    1..461
                     /gene="SPAC9G1.01"
                     /note="nominal overlap with EM_FUN:SPAC3H1 Z68144
                     position36460..36920"
     CDS             <1..476
                     /gene="SPAC9G1.01"
                     /note="SPAC9G1.01, similarity: to YGR066C, YG29_YEAST,
                     P53342, hypothetical 34.0 kd protein, (292aa), fasta
                     scores, opt:210, E():4e-09, (32.7% identity in 159 aa
                     overlap), similarity: to YBR105C, YBV5_YEAST, P38263,
                     hypothetical 41.2 kd protein, (362aa), opt:159,
                     E():1.3e-08, (25.7% identity in 191 aa overlap)"
                     /codon_start=3
```

```
                     /product="hypothetical protein"
                     /protein_id="CAB11487.1"
                     /db_xref="GI:2370550"
                     /db_xref="SWISS-PROT:Q10079"
                     /translation="SMLCGFIHVSTSPNNTSLTTYFAEIIGNRFTFETKWPEWGAS
                     EEIDNRHWKRIGALKSNGKDIPLRLIQPYDPLSRETVYMRWKELAMLDKTVDYQNHNQ
                     SFPFGISYEGFYYISFSQSTGKIKGYYHSEPEKVLFLELNPIIDRTFPVIEFQ"
     gene            1985..6190
                     /gene="wis4"
     CDS             1985..6190
                     /gene="wis4"
                     /note="SPAC9G1.02, len:1401aa"
                     /codon_start=1
                     /product="MAP kinase kinase kinase"
                     /protein_id="CAB11500.1"
                     /db_xref="GI:2370551"
                     /db_xref="SPTREMBL:O14299"
                     /translation="MGLEHTFYPAEDRFEPLLEHSEPVNFVPKENAKSYVRQGFASPH
                     QSLMDNLVDSTESTKRSENFVSHIPLTPSHSGQSEKLMSTRTSHSPYISPTMSYTNHS
                     PANLTRNSSPNHQHYSTTLRSPPSMRGRGIDVNSSHYPHISRPRTSSDSQKMYTRAPV
                     DYYIQENPYFNNIDQDSISDKSLPSTNQSLHHSEEDTESDNDFSESIHPEFDIDVFY
                     KVSNLIYDESDLQDPEKRERLEWHSMLSSVLKGDVMQTEKRRLRLTEPDGHSGTYISE
                     VWLGLQAWLHGRLNADQAEVIRKSREGVVREVIDFQIQDEETTKPPLEQVTELLE
                     KVEQCKQFYISSREMEENVPLSASKEFNYKLNALISWSNVMESIQVETLVLQKWVGND
                     EFDLTMRTQFNYDGVENTSSFVERIFRQSLQRTFEQRTLTTLNRIIHQAKQTISEN
                     AQAFEEMKLPTYEDKLLPLVRFPIKLLEEALRLRLAYAKKIRLAYAKKLIVSSLDFKI
                     ALSVAVRIKREYIKIASPSPGWSLPTNVDEDYDNVLLDSLKFYFKLLTLKLSSGNKNL
                     YFKEIDFLENEWAFLNEHIYWINGDIHMAGQFSYLSNSLLLNVHRYVESHLNGPTER
                     TAASLTNWYSTLLKNTQIRFRKILRFSETLNSRFENASDFVISEGHLPDLVNRLSTG
                     HFLAYTANLERDGVFVIADHTLSENPEALKALLFSKDISNLETIQQNCSYVLILCPVH
                     PIVWKGRIEKVDVPDFSVDLKTNVRIIASNKREHLQAAKSVFQSISGDLVTLAVECR
                     SSITRVYKEFIRLSKLCMRISSTVVDCVSAVREACSGVNCHDLIYHVFSFAAEFGQRI
                     LRFLSFDSYWQTKLKRKITSLAVEWISFICDECDLMDRKTFRWGVGALEFLMLMIRGN
                     NILLIDDAMFLKIREKVGKSMAFLLTHFDVLGAKSKVAAKLQRESTEVSSSPRLTSFG
                     DVEEALSIOLLQKETMLRIDELEIERNNTLLERLAIGHVLDDSVFRNRDFIKLASSF
                     SNIIIRWQQGHFVRSGMFGDVYTGVNEMTGDLLAVKEIKLQDSRTFRSTVDQIHNEMT
                     VLERLNHPNVTYGVEVHREKVYIFMEFCQGGSLADLLAHGRIEDENVLKVYVVQLL
                     EGLAYIHSQHILHRDIKPANILLDHRGMIKYSDFGSALYVSPPTDPEVRYEDIQPELQ
                     HLAGTPMYMAPEILLGTKKGDFGAMDIWSLGCVILEMMTGSTPWSEMDNEWAIMXHVA
                     AMHIPSIPQNEKISSLARDFIEQCFERDPEQRPRAVDLLTHPWITDFRKKTIITMPPA
                     TITKTSLSHTITEEKTAQLLAGRHDDSKAETDSLAASYKEESALPVASNVGLRQPNE
                     LRIDSINLPPAIVTPDTINYSVD"
     misc_feature    5093..5902
                     /gene="wis4"
                     /note="Pfam match to entry PF00069 pkinase, Eukaryotic
                     protein Kinase domain"
     misc_feature    5453..5491
                     /gene="wis4"
                     /note="PS00108 Serine/Threonine protein kinases
                     active-site signature"
     gene            complement(6533..6921)
                     /gene="rpl30"
     CDS             complement(join(6533..6623,6683..6921))
                     /gene="rpl30"
                     /note="SPAC9G1.03c, len:109aa"
                     /codon_start=1
                     /product="60s ribosomal protein L30"
                     /protein_id="CAB11499.1"
                     /db_xref="GI:2240209"
                     /db_xref="SWISS-PROT:P52808"
                     /translation="MASVVTKKSKSGDTINAKLALTMKSGKYVLGYKSTLKTLRSGK
                     AKLILIAGNCPPLRKSELEYAMLSKANVIHYAGTNIDLGTACGKLFRVGVLAITDAG
                     DSDILDA"
     misc_feature    complement(6557..6619)
                     /gene="rpl30"
                     /note="PS00993 Ribosomal protein L30e signature 2"
     misc_feature    complement(6624..6639)
                     /gene="rpl30"
                     /note="ctaactgagttattag, splice branch and acceptor"
     misc_feature    complement(6677..6682)
                     /gene="rpl30"
                     /note="gtaagt, splice donor sequence"
     misc_feature    complement(6772..6846)
```

seq_name: gb_in1:CELF29G9

```
seq_documentation_block:
LOCUS       CELF29G9      42751 bp     DNA            INV      07-AUG-1997
DEFINITION  Caenorhabditis elegans cosmid F29G9.
ACCESSION   AF016440
VERSION     AF016440.1  GI:2315352
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 42751)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 42751)
  AUTHORS   Langston,Y.
  TITLE     The sequence of C. elegans cosmid F29G9
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 42751)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997)
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RQ, England
            e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:

            The 5' cosmid is C10F3, 200 bp overlap;3' cosmid is R01B10, 2200 bp
            overlap. Actual start of this cosmid is at base position 197 of
            CELF29G9; actual end is at 40555 of CELF29G9

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..42751
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
```

```
     promoter        /gene="rpl30"
                     /note="PS00709 Ribosomal protein L30e signature 1"
     promoter        complement(6991..6998)
                     /note="Homol D box"
     gene            complement(7005..7014)
                     /note="Homol E box"
                     8152..9417
                     /gene="oxal"
     CDS             join(8152..8180,8246..8996,9073..9417)
                     /gene="oxal"
                     /note="SPAC9G1.04, len:374aa"
                     /codon_start=1
                     /product="cytochrome oxidase biogenesis protein"
                     /protein_id="CAB11488.1"
                     /db_xref="GI:2370552"
                     /db_xref="SPTREMBL:O14300"
                     /translation="MLHFLIRPSTRCAGSPHFNSTSSFLRKRLQSKTILSKRLGFQFR
                     PSSTNVLAEVSTATSGFNPSWWPYALIQNTAYTINVYAGAPWWVSIILTLGVRLALT
                     PVMIASFRNSTKLSVIQPEMKKELEAIKTAKLDNDQLALNQHSIALRGIYLKHVNPF
                     AIFLPLTQSAVFFSFYAIRKMSRLSVDGFTTGGLAWFKDLSIPDYCLIPTINAGL
                     MFSGMQMNRANTASTIGNSTNWRFFFLCCLLSPLLTAKLPAAIFMYWIPSSLFNIVQ
                     GYILKNPWVRSKLGFAPLPSIIEKQPSGFTLITNPIKSLKEFYKGVRDGFKTQYEQLQ
                     KDVSRRAVATTSASTIRPNSHYKKLKELNRSKKNSKKQSN"
     misc_feature    8181..8186
                     /gene="oxal"
                     /note="gtaagt, splice donor sequence"
     misc_feature    8233..8245
                     /gene="oxal"
                     /note="ttgaccaattag, splice branch and acceptor"
     misc_feature    8997..9002
                     /gene="oxal"
                     /note="gtacgt, splice donor sequence"
     misc_feature    9054..9072
                     /gene="oxal"
                     /note="ctaaccaaagatcaaatag, splice branch and acceptor"
     gene            9769..11928
                     /gene="SPAC9G1.05"
     CDS             join(9769..9873,9935..10058,10129..10230,10271..10316,
                     10423..10468,10518..11459,11506..11928)
                     /gene="SPAC9G1.05"
                     /note="SPAC9G1.05, len:595aa, similarity: to WD42_DICDI,
                     P54686, wd40 repeat protein 2, (597aa), fasta scores,
                     opt:1374, E():0, (36.7% identity in 603 aa overlap), also
                     similar eg. YMR092C, AIP1_YEAST, P46680, actin interacting
                     protein 1, (615aa), fasta scores, opt:94, E():0.fasta
                     scores, opt:1374, E():0, (33.3% identity in 636 aa
                     overlap)"
                     /codon_start=1
                     /product="beta-transducin"
                     /protein_id="CAB11489.1"
                     /db_xref="GI:2370553"
                     /db_xref="SWISS-PROT:O14301"
                     /translation="MSSQLKSTWAPVPSTKPSQPCKIGTDFKGERIVYPANKAIIRE
                     IEKQENCFQYNEHTAPTTVARFSPSGYYVASGDNQGNVRIWDCAGEDKILKNQVTAIS
                     GRITDLDWDGSQRIIAVGEGKERYGHAFTADSGNSVGEIFGHSSVVNAVSLRKSRPF
                     RAATASDDNSINFFHGTPYRFNRSLRVHSKFVYDVRYSPNDERFASAGADGKVVVFDG
                     KTGDQVYEIDAHKGSIFSISWSPDSSQFVTSSAGYSCKIWDANTGSLIREWLSSDKKQ
                     LVGTVWPTKDLIIVVNSKGGNLTYLNPSDCKVIDTIYGHQRSITAATLSPDATHFYTAS
                     YDGTVLSWDIGKQKAFPLVGESHTNQVMQMIMADDHVITIGMDDTLRVIDIKQGCFAK"
```

```
alignment_scores:
       Quality:       45.00          Length:        14
       Ratio:          3.750         Gaps:           0
Percent Similarity: 85.714   Percent Identity: 50.000

alignment_block:
US-08-653-294-21 x SPAC9G1
Align seq 1/1  to: SPAC9G1  from: 1  to: 30985

         2  TyrArgLeuLeuLysValIleArgIleValValLeuLysTyr 15
            ||:::|||||:::::::|:: ::|:||  ||| ||:|| |||||
     12984  TATAAACTGCTGGTCCGGATTCTCAGGTACGTGAGTAAATAT 13025
```

Left column:

```
gene
    /db_xref="taxon:6239"
    /chromosome="V"
    /clone="F29G9"
    13017..15192
    /gene="F29G9"
CDS
    join(13017..13124,13720..14220,14274..14475,15071..15192)
    /gene="F29G9.4"
    /note="Similar to BZIP transcription factor"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB65905.1"
    /db_xref="GI:2315357"
    /translation="MEDDDDDKRLKRRQRNKEAAARCRQRRIDLMKELQDQVNDFKNS
NDKKMAECNNIRNKLNSLKNYLETHDCKLSREERTHEINRLIIPPSTVPPSQPYLQHS
LRVHPPRADSVPYSIRSGHSSSSSEQHSPVEDYKPSIDQLLLPISCIQNIKDRNINS
MPPPALPASTSAAGIHVITSIPVSHANSLHGRSENVFAEPERKIPKIELDQTLTSLTM
PDDVERPSALPTLSRIVENQPITTPSRPFRLGGEYQNQTPQSTGNGLFGGPGPFDLL
SNTGLITPSGQPTMNFVSTPTIPQPHPDADLRPL"
gene
    complement(15701..17169)
    /gene="F29G9.5"
CDS
    complement(join(15701..15925,15972..16694,16740..16960,
17007..17169))
    /gene="F29G9.5"
    /note="strong similarity to the AAA family of ATPases"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB65906.1"
    /db_xref="GI:2315358"
    /translation="MGQQQSGFGGRGNDRGAGDGEKKEKKKYEAPIPSRIGKKKGSK
GPDAASKLPAVTPHARCRLKLLKSERIKDYLLMEQEFIQNOERLKPQEERQEEERAKV
DELRGTPMAVGSLEEIIDDQHAIVSTNVGSEHYVNIMSFVDKEQLEPGCSVLLNHKNH
AVIGVLSDDTDPMVSVMKLEKAPQETYADVGGLDQOIQEIKEAVELPLTHPEYYEMG
IRPPKGVILYGCPGTGKTLLAKAVANOTSATFLRIVGSELIQKYLGDGPKMVRELFRV
AEENAPSIVFIDEIDAVGTKRYDSNSGGEREIQRTMELLNQLDGFDSRGDVKVLMAT
NRIESLDPALIRPGRIDRKIEFPLPDEKTKRRIFQTHTSRMTLGKEVNLEEFITAKDE
LSGADIKAMCTEAGLLALRERRMRVTMEDFQKSKENVLYRKKEGAPEELYL"
gene
    17476..18251
    /gene="F29G9.3"
CDS
    join(17476..17553,17602..17705,17960..18251)
    /gene="F29G9.3"
    /note="Similar to clathrin coat assemble protein; coded
    for by C. elegans cDNA yk175h10.5"
    /codon_start=1
    /protein_id="AAB65902.1"
    /db_xref="GI:2315354"
    /translation="MMQYMLLFSRQGKLRLQKWYTAYPDKQKKKICRELITQILARKP
KMCAFLEYKDLKVVYKRYASLYFCCAIEQNDNELITLEVIHRYVELLDKYPGSVCELD
IIFNFEKAYFLDEFLLAGEIIQETSKKQVLKAIAAQDLIQEEETPQGFFEDHGLG"
gene
    22980..25634
    /gene="F29G9.2"
CDS
    join(22980..23133,23767..23885,23939..24070,24366..24959,
25009..25294,25342..25634)
    /gene="F29G9.2"
    /note="coded for by C. elegans cDNA CEESK57FB"
    /codon_start=1
    /protein_id="AAB65904.1"
    /db_xref="GI:2315356"
    /translation="MSSESGGSASTVHYAKPVLRHVPMPSSTTPSIGSSSSSSYA
SSTKQTPPRSPVIRPTVVVSKNSIATPSSSLTPQGTPSYAVPVSRNQMQYSASKLQY
EHMRHRCKMLDDENQKLMRMQSDVVNDANRRVQMHVNEIRMLKEDNRKLAISNKELRD
RQLCMYLDEQROSLMSNAAANVDCDNESEDLGCGSSEQSGSGSEGHNDEEKHHEFNKCF
NKHKESTLRRIMATSMCSEPSEEERREVSKRERSRLLGIYIQSLENRIKHLEMSQNHE
SFWNSSNVGSDCDEKTIIERGWLGEEVMSNSEDCHLELKPVMTTSSTSSHIFGNDK
CPMFDSMTSNMTSSGCTTYASSGTDGDSFVIGDEIDIGNLEVRTLSRIDEEATSASD
TLKESARMPPKIAPPICSSLVLTNFDNMSEDCAPRLMRSASETCRPTTTLISSTQPAQ
RSVSVEKNNNNVHTHN"
gene
    complement(26014..30269)
    /gene="F29G9.6"
CDS
    complement(join(26014..26095,26138..26278,27281..27361,
27421..27551,28400..28500,28853..28975,29114..29197,
29739..29996,30057..30269))
    /gene="F29G9.6"
```

Right column:

```
    /note="contains similarity to the short chain type of
    alcohol and other dehydrogenases"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB65901.1"
    /db_xref="GI:2315353"
    /translation="MSIPHFLCPFIFSFFPIATISVPILFLFYLSPSLSLLLFRVAH
LLLLLDASKMFEKWAWSPILLFSCSRVTHTHTCSTTTGDPRENLLSSGFHAILKLQ
NLIGRTCQIMTTINYKRTILITGATDGIGKQTALDLAAHPDNFVIIHGRTEEKCIATK
DWIGKENGNCSNIDYVAGDFAVLKEVAIIAEEVERRFPELNVLLCNAGVLYPRRLETK
DGMESTFQVNYLAHYLLCNLLPVLSHNRSNVIVVGSVLHTWPSLDWADVMATKEYEK
YLQYSRSKLMCHLMAFALHRRMNIARQHVNVNIIELGKEKEPNNGKLRTTSALSSSM
STLSICRQAGNLAQLIEGPCLEKISGKYLDPSGKQMRSGSDATDERLQERLWAYSKEL
CHEYL"
gene
    complement(33753..36485)
    /gene="F29G9.7"
CDS
    complement(join(33753..33916,34076..34173,34222..34536,
35131..35405,35616..35934,35983..36061,36467..36485))
    /gene="F29G9.7"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB65907.1"
    /db_xref="GI:2315359"
    /translation="MPEYVPTLNFNSTSSIFKRSVKFGDLRLPIKYDPETAYTCST
DNYEDCLQPCLDLFATIFSCRGGIAIVNLNECSLNVREHTAHRIFNKNLVVKLCGDVQ
NLGTNNGSLQSFYPEYVLLEGIVADFHFLEEEFDPYEMFTRNRVSIMDAKWVTLYHLK
SLKFIWLVLYNSMLTERDVNEFIKLWLDGKWLQLHREFEIRHTSRGFGFDLEVLKSGLNV
KKWNKNLRNGCYIVLASLESIISKKTISSSKMRRIVRARALLFLGERSDHALSTQPPP
PIKFVILSYSPSQTVQWKPSVVSSVICRKRQEQPTFERLLQLPPAFANILKILGAV
KILELSQLSKRMCTLIQAANTFDHIMIAFSESSALPIIPKSVAMLQTLNLAYQPAAG
ISSTFSCIMCRAPQNGELSLNKTILITLTLW"
gene
    39027..40103
    /gene="F29G9.1"
CDS
    join(39027..39211,39749..40103)
    /gene="F29G9.1"
    /codon_start=1
    /evidence=not_experimental
    /protein_id="AAB65903.1"
    /db_xref="GI:2315355"
    /translation="MRTDVGPLRYDRAVLALWNTWVMMCSSYGVEHHMKQLLFKAI
VTTCIIRKQQQGSAKGRPTEEEGTNDDGKKTWTKGGDSMELWNEISLKRSRATMDRD
GSGWNDRFECDNRDRRTGWIRDGFGMDSGWIGMDRDGLGWIGMDRDGSEWVAMDRDGS
GWVGKDRDGSVWIEMDGIG"

BASE COUNT     13491 a      7983 c      7355 g     13922 t
ORIGIN

alignment_scores:
            Quality:    45.00           Length:        14
            Ratio:       3.462          Gaps:           0
    Percent Similarity: 92.857     Percent Identity: 57.143

alignment_block:
US-08-653-294-21 x CELF29G9

Align seg 1/1   to: CELF29G9   from: 1   to: 42751

        2 TyrArgLeuLeuIleLysValIleArgIleValLeuLysTyr  15
          |||:||||||||||:::|||:::|||:|||   |||||||||
    26955 TACAATGTTTAATACGAATTATTAAAATAAAATTAAAGTAT  26996
```

OM of: US-08-653-294-21 to: N_Geneseq_36:*   out_format : pfs

Date: Feb 8, 2000 7:30 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_l/USPTO_spool/US08653294/rrunat_04022000_160701_15807/app_query.fasta.2
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEX=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US08653294 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-08-653-294-21
Query length: 15
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 873.190000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X13927	+	42.00	125.48	21.76	810	! H. pylori GHPO 635 gene. New is
N_Geneseq_36:Q56758	+	42.00	116.44	69.39	2255	! Chitinase derivative gene #1.
N_Geneseq_36:X13348	+	41.00	108.74	186.35	3721	! Enterococcus faecalis genome c
N_Geneseq_36:T33641	+	40.00	118.30	54.67	869	! Aspergillus arabinofuranosidase
N_Geneseq_36:T33646	+	40.00	108.78	185.41	2555	! Aspergillus arabinofuranosidas
N_Geneseq_36:X13195	+	40.00	97.67	770.82	8991	! Enterococcus faecalis genome c
N_Geneseq_36:X30892	+	39.00	118.40	54.01	593	! Streptococcus pneumoniae genomi
N_Geneseq_36:X42962	+	39.00	107.77	210.96	1975	! Streptococcus pneumoniae polyp
N_Geneseq_36:V07122	+	39.00	105.34	288.02	2600	! Cytophaga drobachiensis kappa-
N_Geneseq_36:T84234	+	39.00	103.76	352.79	3110	! DNA encoding an autolysin and
N_Geneseq_36:V53557	+	39.00	103.76	352.79	3110	! DNA encoding 2 Staphylococcus
N_Geneseq_36:V52309	+	39.00	101.58	467.01	3984	! Streptococcus pneumoniae genom
N_Geneseq_36:V69717	+	39.00	101.47	473.26	4031	! Tumour rejection antigen precu
N_Geneseq_36:V69720	+	39.00	101.06	499.13	4225	! Tumour rejection antigen precu
N_Geneseq_36:V74331	+	39.00	93.04	1.4e+03	10470	! Staphylococcus aureus contig
N_Geneseq_36:V70401	+	39.00	89.60	2.2e+03	15462	! Human parainfluenza virus typ
N_Geneseq_36:X20248_04	+	39.00	72.28	2.0e+04	110000	! Continuation (5 of 10) of
N_Geneseq_36:X20248_04	-	39.00	72.28	2.0e+04	110000	! Continuation (5 of 10) of
N_Geneseq_36:X20248_08	+	39.00	72.28	2.0e+04	110000	! Continuation (9 of 10) of
N_Geneseq_36:Q03147	+	38.00	105.55	280.35	1751	! Yeast Ubiquitin hydrolase gene
N_Geneseq_36:X02969	+	38.00	105.40	285.79	1781	! Human IL-1ra BAC contiguous DN
N_Geneseq_36:X03047	+	38.00	90.95	1.8e+03	9158	! Human IL-1ra BAC contiguous DN
N_Geneseq_36:V74364	+	38.00	85.70	3.6e+03	16592	! Staphylococcus aureus contig
N_Geneseq_36:V52273	-	38.00	80.80	6.7e+03	28882	! Streptococcus pneumoniae geno
N_Geneseq_36:T58840_4	+	38.00	69.00	3.0e+04	110000	! Continuation (5 of 6) of T5
N_Geneseq_36:V21209_10	-	38.00	69.00	3.0e+04	110000	! Continuation (11 of 17) of
N_Geneseq_36:X20248_06	+	38.00	69.00	3.0e+04	110000	! Continuation (7 of 10) of
N_Geneseq_36:X20250	+	38.00	68.89	3.0e+04	111309	! Borrelia burgdorferi polynuc
N_Geneseq_36:V74525	+	37.50	90.76	1.9e+03	7769	! Staphylococcus aureus contig
N_Geneseq_36:X00666	-	37.00	107.12	229.22	1011	! Human secreted protein gene 56
N_Geneseq_36:V69759	+	37.00	106.78	239.52	1051	! Arabidopsis ovary-specific AGI
N_Geneseq_36:V29580	-	37.00	105.71	274.65	1186	! Pythium oligandrum isolate 23-
N_Geneseq_36:X06780	+	37.00	102.29	426.14	1748	! Human testis secreted protein
N_Geneseq_36:X13220	-	37.00	96.20	930.92	3485	! Enterococcus faecalis genome c
N_Geneseq_36:X13280	-	37.00	95.19	1.1e+03	3907	! Enterococcus faecalis genome c
N_Geneseq_36:V52207	-	37.00	82.45	5.4e+03	16535	! Streptococcus pneumoniae geno
N_Geneseq_36:X13336	+	37.00	76.41	1.2e+04	32768	! Enterococcus faecalis genome
N_Geneseq_36:V30458_3	-	37.00	65.72	4.5e+04	110000	! Continuation (4 of 6) of V3
N_Geneseq_36:V21209_01	-	37.00	65.72	4.5e+04	110000	! Continuation (2 of 17) of
N_Geneseq_36:X20248_08	-	37.00	65.72	4.5e+04	110000	! Continuation (9 of 10) of
N_Geneseq_36:V30459_3	-	37.00	65.72	4.5e+04	110000	! Continuation (4 of 6) of V3

N_Geneseq_36:Q46292	-	36.50	101.95	444.90	1508	! PAL regulator gene. Regulat
N_Geneseq_36:Q46293	-	36.50	92.47	1.5e+03	4415	! Phenylalanine ammonia lyase
N_Geneseq_36:T78725	-	36.00	127.42	16.97	70	! SELEX generated ligand to hCG
N_Geneseq_36:V89042	+	36.00	114.00	94.88	320	! EST clone BV31. New polynucl

seq_name: N_Geneseq_36:X13927

seq_documentation_block:
ID X13927 standard; DNA; 810 BP.
AC X13927;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 635 gene.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key   Location/Qualifiers
FT CDS   98..760
FT   /*tag= a
PN WO984347B-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR P-PSDB; W98208.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 111-113; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases. e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 810 BP; 223 A; 123 C; 188 G; 276 T;

alignment_scores:
        Quality: 42.00   Length: 15
          Ratio: 3.231   Gaps: 0
Percent Similarity: 86.667   Percent Identity: 53.333

alignment_block:
US-08-653-294-21 x X13927

Align seg 1/1 to: X13927 from: 1 to: 810

   1 AlaTyrArgLeuLeuLysValIleArgValLeuLysTyr 15
     ||||||| |||||||::||||||| |||::::::::||
 720 GCTTACCGTAGCTTATTTCAATTAATCTATAATTTTAGATAT 764

seq_name: N_Geneseq_36:Q56758

seq_documentation_block:
ID Q56758 standard; DNA; 2255 BP.
AC Q56758;
DT 12-OCT-1994 (first entry)
DE Chitinase derivative gene #1.
KW Chitinase; derivative; beta-1,4-glycoside bond; chitin;
KW beta-N-acetylhexosaminidase activity; yeast; ss.
OS Rhizopus niveus.
FH Key   Location/Qualifiers
FT cds   138..1750
FT   /*tag= a
FT   /product= Chitinase
FT exon   138..242

```
FT   intron          /*tag= b
FT                   243..305
FT   exon            /*tag= c
FT                   306..498
FT   intron          /*tag= d
FT                   499..566
FT   exon            /*tag= e
FT                   567..1750
FT                   /*tag= f
PN   J06046849-A.
PD   22-FEB-1994.
PF   28-JUL-1992; 201427.
PR   28-JUL-1992; JP-201427.
PA   (KAGO ) KAGOME KK.
DR   WPI; 94-097015/12.
DR   P-PSDB; R48669.
PT   New chitinase enzyme and coding sequence - cleaves beta-1,4-glycoside
PT   bond of chitin but has no beta-N-acetylhexosaminidase activity
PS   Claim 11; Page 21-23; 66pp; Japanese.
CC   The sequences given in Q56756-61 encode chitinase derivatives which
CC   cleave the beta-1,4-glycoside bond of chitin but have substantially
CC   no beta-N-acetylhexosaminidase activity. These sequences may be
CC   introduced into yeast and cultured for the production of the
CC   chitinase proteins.
SQ   Sequence  2255 BP;  614 A;  448 C;  446 G;  747 T;

alignment_scores:
     Quality:  42.00          Length:       9
     Ratio:     4.667         Gaps:         0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-08-653-294-21 x Q56758

Align seg 1/1  to Q56758  from: 1  to: 2255

   7 LysValIleArgIleValLeuLysTyr 15
     |||||||||||||||||||||||||||
1867 AAAGTAATAAGAATTGTCTTAAAATAT 1893

seq_name: N_Geneseq_36:X13348

seq_documentation_block:
ID   X13348 standard; DNA; 3721 BP.
AC   X13348;
DT   19-MAR-1999  (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO:411.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; U08985.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 1626-1628; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence  3721 BP;  1387 A;  582 C;  785 G;  954 T;

alignment_scores:
     Quality:  41.00          Length:      14
     Ratio:     3.417         Gaps:         0
Percent Similarity: 85.714    Percent Identity: 50.000

alignment_block:
US-08-653-294-21 x X13348/rev

Align seg 1/1  to reverse of: X13348  from: 1  to: 3721

  2 TyrArgLeuGluLeuLysValIleArgIleValLeuLysTyr 15
    ||||| ::::::: ||:::|||||||||||||:::::::::
698 TACCGATCCATGCTTAAATTGATCAGAATTATTTTCGGATAC 657

seq_name: N_Geneseq_36:T33641

seq_documentation_block:
ID   T33641 standard; DNA; 869 BP.
AC   T33641;
DT   11-DEC-1996  (first entry)
DE   Aspergillus arabinofuranosidase gene promoter.
KW   Promoter; arabinofuranosidase; AbfC; arabinoxylan;
KW   viscosity modifier; food; feedstuff; ss.
OS   Aspergillus niger strain 3M43.
PN   WO9629416-A1.
PD   26-SEP-1996.
PF   11-MAR-1996; E01009.
PR   17-MAR-1995; GB-005479.
PA   (DANI-) DANISCO AS.
PI   Baruch A, Madrid SM,  Rasmussen P;
DR   WPI; 96-443191/44.
PT   Aspergillus arabinofuranosidase - useful for degradation of
PT   arabinoxylan
PS   Claim 8; Page 48; 105pp; English.
CC   The promoter (T33641) of the arabinofuranosidase gene of Aspergillus
CC   niger 3M43 was identified in a full-length gene (T33646) isolated
CC   from a genomic DNA library. The promoter is strongly repressed by
CC   glucose and induced by intermediates of xylose metabolism. It
CC   can be used to control the expression of a gene of interest, e.g.
CC   A. niger arabinofuranosidase or a heterologous gene, in a host
CC   cell, pref. Aspergillus or a transgenic plant.
SQ   Sequence  869 BP;  224 A;  209 C;  193 G;  243 T;

alignment_scores:
     Quality:  40.00          Length:      12
     Ratio:     3.333         Gaps:         0
Percent Similarity: 100.000   Percent Identity: 58.333

alignment_block:
US-08-653-294-21 x T33641

Align seg 1/1  to T33641  from: 1  to: 869

  2 TyrArgLeuGluLeuLysValIleArgIleValLeu 13
    |||::|||||||::||||||||||||||:::::::::
165 TACCAACTTTAGTTCGATTGATCAGAATACTTTG 200

seq_name: N_Geneseq_36:T33646

seq_documentation_block:
ID   T33646 standard; DNA; 2555 BP.
AC   T33646;
DT   11-DEC-1996  (first entry)
```

DE   Aspergillus arabinofuranosidase gene.
KW   Arabinofuranosidase; AbfC; arabinoxylan; viscosity modifier; food;
KW   feedstuff; ds.
OS   Aspergillus niger strain 3M43.
FH   Key   Location/Qualifiers
FT   promoter   1..869
FT   /*tag= a
FT   cds   870..1757
FT   /*tag= b
FT   signal_peptide 870..947
FT   /*tag= c
FT   mat_peptide   948..1754
FT   /*tag= d
FT   terminator   1755..2555
FT   /*tag= e
PN   WO9629416-A1.
PD   26-SEP-1996.
PF   11-MAR-1996; E01009.
PR   17-MAR-1995; GB-005479.   --
PA   (DANI-) DANISCO AS.
PI   Baruch A, Madrid SM, Rasmussen P;
DR   WPI; 95-443191/44.
DR   P-PSDB; W00810.
PT   Aspergillus arabinofuranosidase - useful for degradation of
PT   arabinoxylan
PS   Example; Page 52-55; 105pp; English.
CC   The Aspergillus niger 3M43 arabinofuranosidase gene (T33646)
CC   codes for the precursor form (W00810) of the arabinofuranosidase
CC   enzyme. It was isolated from a genomic DNA library by screening
CC   with a PCR clone obtd. by amplification of A. niger DNA using
CC   primers (see also T33644-45) based on an isolated peptide (W00806)
CC   the enzyme. The mature enzyme coding sequence (see also T33640)
CC   can be used for prodn. of arabinofuranosidase (W04167) in
CC   transformed hosts, pref. Aspergillus or a transgenic plant.   The
CC   promoter (see also T33641), terminator (T33642) and signal sequence
CC   (T33643) may also be used to control expression of the
CC   arabinofuranosidase gene, or other gene of interest, in a host
CC   cell.
SQ   Sequence   2555 BP;   632 A;   651 C;   615 G;   657 T;

alignment_scores:
         Quality: 40.00   Length: 12
         Ratio: 3.333   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 58.333

alignment_block:
US-08-653-294-21 x T33646   ..

Align seg 1/1 to: T33646   from: 1 to: 2555

        2 TyrArgLeuLeuIleLysValIleArgIleValLeu 13
          |||::||||||:::::::::|||||||:::|||
      165 TACCAACTTTAGTTCGATTGATCAGAATACTTTTG 200

seq_name: N_Geneseq_36-X13195

seq_documentation_block:
ID   X13195 standard; DNA; 8991 BP.
AC   X13195;
DT   19-MAR-1999 (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO:258.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; U08985.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.   --
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;

DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 1239-1244; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence   8991 BP;   2910 A;   1478 C;   1913 G;   2683 T;

alignment_scores:
         Quality: 40.00   Length: 13
         Ratio: 3.333   Gaps: 0
Percent Similarity: 92.308   Percent Identity: 53.846

alignment_block:
US-08-653-294-21 x X13195/rev

Align seg 1/1 to reverse of: X13195   from: 1 to: 8991

        2 TyrArgLeuLeuIleLysValIleArgIleValLeuLeuLys 14
          ::: ||||||||||::||||||||||||||:|||||||:::
     1946 TTCGCCTCCTCATCAAAATAGTACAGATAGTCCTGCGC 1908

seq_name: N_Geneseq_36-X30892

seq_documentation_block:
ID   X30892 standard; DNA; 593 BP.
AC   X30892;
DT   20-MAY-1999 (first entry)
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:169.
KW   Streptococcus pneumoniae strain O100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
OS   Streptococcus pneumoniae.
PN   WO9737026-A1.
PD   09-OCT-1997.
PF   01-APR-1997; U05306.
PR   22-AUG-1996; US-025788.
PR   02-APR-1996; US-014690.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Stodola RK;
DR   WPI; 97-503111/46.
DR   P-PSDB; Y11310.
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
PS   Claim 5; Page 156; 354pp; English.
CC   X30724 to X30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain O100993. These genomic DNA sequences
CC   encode the novel proteins given in Y11114 to Y11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
CC   be used to treat an individual having need to inhibit a bacterial
CC   protein. Vectors expressing the proteins can be used to induce a
CC   protective immune response in mammals.
SQ   Sequence   593 BP;   206 A;   76 C;   117 G;   194 T;

```
alignment_scores:
    Quality:            39.00              Length:          14
    Ratio:              3.250              Gaps:             0
Percent Similarity:    85.714     Percent Identity:    50.000

alignment_block:
US-08-653-294-21 x X30892/rev         ..
Align seg 1/1  to reverse of: X30892  from: 1  to: 593

    2 TyrArgLeuIleLysValIleArgIleValLeuLysTyr 15
      |||:::|||:::  |||:::|||:::|||||:::||||
   69 TACCAACTTGTGAGTATGGTTTGAGAATGGTACTACGTTAC 28

seq_name: N_Geneseq_36:V42962

seq_documentation_block:
ID  V42962 standard; DNA; 1975 BP.
AC  V42962;
DT  09-NOV-1998  (first entry)
DE  Streptococcus pneumoniae polypeptide coding region.
KW  Polypeptide: ORF; open reading frame; infection; bacterial;
KW  streptococcal; bacteremia; diagnosis; prophylaxis; ds.
OS  Streptococcus pneumoniae.
FH  Key             Location/Qualifiers
FT  CDS             complement (1724..1888)
FT                  /*tag= a
FT                  /note= "polypeptide"
PN  WO9823631-A1.
PD  04-JUN-1998.
PF  24-NOV-1997; U21976.
PR  27-NOV-1996; US-031879.
PA  (SMIK ) SMITHKLINE BEECHAM CORP.
PA  (SMIK ) SMITHKLINE BEECHAM PLC.
PI  Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI  Reid RH, Zarfos PN;
DR  WPI; 98-322654/28.
DR  P-PSDB: W62682.
PT  Streptococcus pneumoniae polynucleotides - useful for developing
PT  products for diagnosis, prevention and treatment of infections e.g.
PT  pneumonia, bacteremia, meningitis or endocarditis
PS  Claim 1; Page 55; 181pp; English.
CC  The sequence is that of a Streptococcal polypeptide coding region.
CC  The polypeptide can potentially be used for the diagnosis and
CC  prevention of bacterial infections, especially SP infection.
CC  It may be used for the treatment of diseases such as otitis media,
CC  conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC  empyema, endocarditis or infection of the cerebrospinal fluid.
SQ  Sequence 1975 BP;  517 A;  524 C;  379 G;  555 T;

alignment_scores:
    Quality:            39.00              Length:          12
    Ratio:              3.545              Gaps:             0
Percent Similarity:    91.667     Percent Identity:    66.667

alignment_block:
US-08-653-294-21 x V42962              ..
Align seg 1/1  to: V42962  from: 1  to: 1975

    3 ArgLeuLeuLysValIleArgIleValLeuLeuLys 14
      |||||||   :::::||||||||||||||||||
  959 CGCTTGCTCTGTCAAATCATGCGTATAGTCCTCAAG 994

seq_name: N_Geneseq_36:V07122

seq_documentation_block:
ID  V07122 standard; DNA; 2600 BP.
AC  V07122;
DT  29-SEP-1998  (first entry)
DE  Cytophaga drobachiensis kappa-carrageenase coding sequence.
KW  cgkA; carrageenase; carrageenan; D-galactose; glycosidic linkage; ss;
KW  ester-sulphate; disaccharide; glycosyl hydrolase; immunomodulator;
KW  hydrophobic cluster analysis; antiviral agent; growth factor; growth;
KW  keratinocyte proliferation; fibroblast contraction; adhesion; plant.
OS  Cytophaga drobachiensis.
FH  Key             Location/Qualifiers
FT  CDS             875..2512
FT                  /*tag= a
FT                  /gene= "cgkA"
FT                  /product= kappa_carrageenase
FT  sig_peptide     875..979
FT                  /*tag= b
FT                  /note= "putative signal sequence"
FT  mat_peptide     980..2509
FT                  /*tag= c
PN  FR2754271-A1.
PD  10-APR-1998.
PF  07-OCT-1996; 012204.
PR  07-OCT-1996; FR-012204.
PR  03-FEB-1997; FR-001148.
PA  (GOEM-) LAB GOEMAR SA.
PI  Barbeyron T, Henrissat B, Kloareg B, Potin P, Richard C;
DR  WPI; 98-233084/21.
DR  P-PSDB; W50908.
PT  Genes coding for glycosyl hydrolases - useful for producing
PT  recombinant glycosyl hydrolases, especially Cytophaga drobachiensis
PT  kappa-carrageenase
PS  Example; Page 20-23; 31pp; French.
CC  This sequence represents the novel cgkA gene from the microorganism
CC  Cytophaga drobachiensis that encodes the enzyme kappa-carrageenase.
CC  The sequence was isolated from a C. drobachiensis genomic library based
CC  on growth activity on medium containing carrageenan.  From the screens of
CC  the library, 5 final positive clones were isolated, each contained the
CC  present sequence as an insert.  The enzyme acts on kappa-carrageenans
CC  (polymers of D-galactose linked by beta(1-4) and alpha(1-3) glycosidic
CC  linkages and having 1 ester-sulphate groups per disaccharide unit) to
CC  produce oligo-carrageenans.  The invention relates to the isolation of
CC  genes coding for glycosyl hydrolases, which, when subjected to
CC  hydrophobic cluster analysis (HCA) together with Alteromonas
CC  carrageenovora kappa-carrageenase, give an HCA score of at least 75% in
CC  the domain between positions 117 and 262 of the kappa-carrageenase amino
CC  acid sequence.  The enzymes can be used to convert carrageenans to
CC  sulphated oligocarrageenans which are immunomodulators, antiviral agents
CC  and growth factor binding antagonists.  Oligocarrageenans have an
CC  (unspecified) action on keratinocyte proliferation, fibroblast
CC  contractile power and adhesion, growth and protein synthesis in human
CC  cell cultures, and elicit enzyme markers of growth or phenolic defence
CC  metabolism in plants.
SQ  Sequence 2600 BP;  874 A;  541 C;  522 G;  663 T;

alignment_scores:
    Quality:            39.00              Length:          13
    Ratio:              3.250              Gaps:             0
Percent Similarity:    92.308     Percent Identity:    46.154

alignment_block:
US-08-653-294-21 x V07122
Align seg 1/1  to: V07122  from: 1  to: 2600

    2 TyrArgLeuIleLysValIleArgIleValLeuLys 14
      |||:::| |||:::||:::||::::||:::|||:::|
  764 TATAAGAAATTACTAAAATTAGTTAGGATATACTACAA 802

seq_name: N_Geneseq_36:T84234

seq_documentation_block:
ID  T84234 standard; DNA; 3110 BP.
AC  T84234;
DT  01-SEP-1998  (first entry)
DE  DNA encoding an autolysin and 3 unknown proteins.
KW  Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW  Staphylococcal gene; regulatory element; bacterial gene expression;
```

KW   vaccine: Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome; ss.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   CDS   679..1272
FT   /*tag= a
FT   CDS   1668..1844
FT   /*tag= b
FT   CDS   1885..2064
FT   /*tag= c
FT   CDS   2326..2832
FT   /*tag= d
FT   /product= "autolysin"
PN   WO9730070-A1.
PD   21-AUG-1997.
PF   19-FEB-1997; U02318.
PR   20-FEB-1996; US-011888.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR   WPI: 97-424969/39.
DR   P-PSDB: W28340, W28341, W28342, W28342, W28343.
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT   to isolate antimicrobial compounds, and in vaccines against S.
PT   aureus infection
PS   Claim 9; Page 977-978; 989pp; English.
CC   The present sequence encodes 3 Staphylococcus aureus proteins of
CC   unknown function and a protein, that, based on homology with a
CC   known Staphylococcus aureus protein, is believed to be an autolysin
CC   (EC 3.5.1.28) (N-acetylmuramoyl-L-alanine amidase).
CC   The present sequence was obtained from a
CC   library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC   sequence can be used in the construction of ribozymes and antisense
CC   sequences to control the expression of Staphylococcal genes. The DNA
CC   sequence is also useful as a source of regulatory elements for the
CC   control of bacterial gene expression. The encoded protein may be used
CC   to produce vaccines to enable a host to produce specific antibodies
CC   with antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
SQ   Sequence 3110 BP;   488 C;   1111 A;   675 G;   833 T;

alignment_scores:
         Quality: 39.00   Length: 14
         Ratio: 3.250   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 35.714

alignment_block:
US-08-653-294-21 x T84234

Align seg 1/1 to: T84234   from: 1 to 3110

    2 TyrArgLeuLeuIleLysValIleArgIleValLeuLysTyr 15
       ||||::::::   :::||| ||||||:||::::::: :::|||
  785 TACCAGGTAGCGTTGAAAGTTATGAGGGTTATTATCGGTAT 826

seq_name: N_Geneseq_36:v53557

seq_documentation_block:
ID   V53557 standard; DNA; 3110 BP.
AC   V53557;
DT   30-OCT-1998, (first entry)
DE   DNA encoding 2 Staphylococcus aureus proteins of unknown function.
KW   Staphylococcus aureus protein; immune response induction; eye infection;
KW   antibody production; T-cell immune response; gastrointestinal infection;
KW   respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW   central nervous system; kidney infection; urinary tract infection;
KW   antimicrobial compound identification; broad spectrum antibiotic;
KW   therapy; ss.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers

FT   CDS   1208..1300
FT   /*tag= a
FT   /note= "encoded protein shown in W77771"
FT   CDS   1282..1668
FT   /*tag= b
FT   /note= "encoded protein shown in W77772"
PN   EP-841394-A2.
PD   13-MAY-1998.
PF   24-SEP-1997; US-027032.
PR   24-SEP-1996; US-027032.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
PI   Black MT, Burnham MKR, Hodgson JE, Knowles DJC,
PI   Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
PI   Ward JM;
DR   WPI: 98-252940/23.
DR   P-PSDB: W77771, W77772.
PT   New nucleic acid sequences from Staphylococcus aureus WCHU29 -
PT   useful in vaccines and for treatment of bacterial infections of e.g.
PT   respiratory tract and central nervous system
PS   Claim 1: Page 214-216; 390pp; English.
CC   This sequence encodes 2 Staphylococcus aureus proteins of unknown
CC   function, and represents a DNA Sequence of the invention.
CC   The DNA sequences were isolated from Staphylococcus aureus WCHU29
CC   (NCIMB 40771). Host cells containing the DNA sequences are used to
CC   produce polypeptides or fragments. The proteins are used in the treatment
CC   of disease, for inducing an immune response by administering them, to
CC   produce antibody and/or T-cell immune response. Antagonists of the
CC   proteins are used for the inhibition of bacterial polypeptides.
CC   Conditions which may be treated include bacterial infections, especially
CC   respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC   urinary tract, skin, bones and joints. The proteins can also be used to
CC   identify antimicrobial compounds which are broad spectrum antibiotics,
CC   especially useful in the treatment of H. pylori infection.
SQ   Sequence 3110 BP;   488 C;   1111 A;   676 G;   832 T;

alignment_scores:
         Quality: 39.00   Length: 14
         Ratio: 3.250   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 35.714

alignment_block:
US-08-653-294-21 x V53557

Align seg 1/1 to: V53557   from: 1 to: 3110

    2 TyrArgLeuLeuIleLysValIleArgIleValLeuLysTyr 15
       ||||::::::   :::||| ||||||:||::::::: :::|||
  785 TACCAGGTAGCGTTGAAAGTTATGAGGGTTATTATCGGTAT 826

seq_name: N_Geneseq_36:v52309

seq_documentation_block:
ID   V52309 standard; DNA; 3984 BP.
AC   V52309;
DT   23-OCT-1998 (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:176.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay.
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI: 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae

PS Claim 1; Page 1108-1110; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 3984 BP; 1241 A; 867 C; 662 G; 1214 T;

alignment_scores:
       Quality: 39.00   Length: 14
       Ratio: 3.250   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 50.000

alignment_block:
US-08-653-294-21 x V52309

Align seg 1/1 to: V52309 from: 1 to: 3984

      2 TyrArgLeuLeuLeuLysValIleArgIleValLeuLysTyr 15
        |||:|||:||: |||:|||::||||::|||||::|:|||
   2372 TACCAACTTGTGAGTATGGTTTGAGAATGGTACTACGTTAC 2413

seq_name: N_Geneseq_36:V69717

seq_documentation_block:
ID V69717 standard; cDNA; 4031 BP.
AC V69717;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   257..3475
FT   /*tag= a
PN WO9849184-A1.
PD 05-NOV-1998.
PF 24-APR-1998; U08493.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, De Smet C, Lucas S;
DR WPI; 99-024041/02.
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 26; Page 40-42; 84pp; English.
CC This nucleotide sequence comprises novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 cDNA (see also V69720). MAGE-C1
CC is a novel member of the MAGE family that may be recognised by
CC cytoxic T cells, leading to lysis of the tumour cells which
CC express it. MACE-C1 and MACE-C2 (see W81546-47) are expressed in a
CC variety of tumours and in normal testis cells, but not by other
CC normal cells. The MAGE-C1 cDNA was isolated from a melanoma
CC LB373-MEL cDNA library using a probe generated from LB-373-MEL
CC RNA by PCR (see V69728-29). It shows homology to MAGE-A1 cDNA (see

CC V69719) and codes for a putative 1072-amino acid protein. The
CC MACE-C1 gene was localised to chromosome region Xq26-q27. MACE-C1
CC and MACE-C2 cDNAs (see V69726) are claimed, as are: expression
CC vectors; transformed or transfected cell lines (e.g. COS and CHO);
CC an isolated TRAP encoded by the cDNAs; a kit useful in a PCR based
CC assay comprising an oligonucleotide having a sequence of
CC nucleotides 18-34 of the 4031 bp MACE-C1 cDNA and an oligonucleotide
CC having a sequence which is complementary to nucleotides 200-217 of
CC the 4031 bp cDNA sequence; a method for determining expression of
CC a MACE-C1 gene using the kit; a polytope comprising a number of
CC tumour rejection antigens derived from MAGE-C1 or MAGE-C2; and a
CC polytope comprising at least one tumour rejection antigen derived
CC from MAGE-C1 or MAGE-C2 and at least one other tumour rejection
CC antigen. MACE-C1 and MACE-C2 can be used in a method for determining
CC the presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
SQ Sequence 4031 BP; 849 A; 1138 C; 877 G; 1167 T;

alignment_scores:
       Quality: 39.00   Length: 13
       Ratio: 3.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 38.462

alignment_block:
US-08-653-294-21 x V69717

Align seg 1/1 to: V69717 from: 1 to: 4031

      2 TyrArgLeuLeuIleLysValIleArgIleValLeuLys 14
        |||||:::||:||:|||::||||:::|:|:|:||::|||
   1317 TACAGAGTTCTCCTGAGAGTGCTCAAAGTGCTTTTGAGG 1355

seq_name: N_Geneseq_36:V69720

seq_documentation_block:
ID V69720 standard; cDNA; 4225 BP.
AC V69720;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   257..3685
FT   /*tag= a
PN WO9849184-A1.
PD 05-NOV-1998.
PF 24-APR-1998; U08493.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, De Smet C, Lucas S;
DR WPI; 99-024041/02.
DR P-PSDB; W81546.
PT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 1; Page 48-50; 84pp; English.
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 (see W81546). MAGE-C1 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MACE-C1 and MACE-C2 (see W81547) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C1 cDNA was isolated from a melanoma LB373-MEL cDNA by
CC PCR amplification (see V69732-33). It shows homology to MAGE-A1
CC cDNA (see V69719). The MACE-C1 gene was localised to Xq26-q27.
CC MACE-C1 and MACE-C2 cDNAs (see V69726) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MACE-C1
CC gene using the kit; a polytope comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polytope

CC   comprising at least one tumour rejection antigen derived from
CC   MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC   MACE-C1 and MACE-C2 can be used in a method for determining the
CC   presence of cytolytic T cells specific for complexes of a human
CC   leukocyte antigen (HLA).
SQ   Sequence   4225 BP;   871 A;   1198 C;   923 G;   1233 T;

alignment_scores:
        Quality: 39.00   Length: 13
        Ratio: 3.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 38.462

alignment_block:
US-08-653-294-21 x V69720

Align seg 1/1 to: V69720 from: 1 to: 4225

    2 TyrArgLeuLeuIleLysValIleArgIleValLeuLys 14
      |||||:||::||||||||:|:||::|||:|||||:
 1527 TACAGAGTTCTCCTGAGAGTGCTCAAAGTGCTTTTGAGG 1565

seq_name: N_Geneseq_36:V74331

seq_documentation_block:
ID   V74331 standard; DNA; 10470 BP.
AC   V74331;
DT   16-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #20.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
OS   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   misc_feature   361..420
FT   /*tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   2161..2220
FT   /*tag= b
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   3961..4020
FT   /*tag= c
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   5761..5820
FT   /*tag= d
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   7561..7620
FT   /*tag= e
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT   misc_feature   9361..9420
FT   /*tag= f
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.

PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 271-277; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   10470 BP;   2804 A;   2009 C;   1550 G;   3742 T;

alignment_scores:
        Quality: 39.00   Length: 14
        Ratio: 3.250   Gaps: 0
Percent Similarity: 85.714   Percent Identity: 35.714

alignment_block:
US-08-653-294-21 x V74331/rev

Align seg 1/1 to reverse of: V74331 from: 1 to: 10470

    2 TyrArgLeuLeuIleLysValIleArgIleValLeuLysTyr 15
      |||:::::: :::||||||||||||:::::::: :::|||
 9287 TACCAGGTAGCGTTGAAAGTTATGAGGGTTATTTATCGGTAT 9246

THIS PAGE BLANK (USPTO)